

BAYESIAN APPROACH TO THE ADAPTIVE DPCM FOR THE $AR(P)$ AND
TWO DIMENSIONAL $AR(1,1)$ MODEL

By

FANGSHI SUN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1992

TO MY PARENTS AND MY WIFE

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Mark Yang for his guidance and support during my years at University of Florida. He suggested this topic and was a major influence in its development. I would wish to thank Dr. Rosalsky, Dr. Ballerini and Dr. Miller, members of my supervisory committee, for their advice and many helpful suggestions. I also wish to acknowledge valuable help from Dr. Chang.

Further, I would like to thank the entire Department of Statistics for the encouragement, support and friendships they have provided.

TABLE OF CONTENTS

	<u>PAGE</u>
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
ABSTRACT	vii
 CHAPTERS	
1 INTRODUCTION	1
1.1 Quantization of Continuous Signals	1
1.2 Classification of the Quantization Schemes	3
1.3 The Bayesian Approach	4
2 QUANTIZATION SCHEMES	8
2.1 Introduction	8
2.2 Minimum Mean Square Error Quantization	9
2.3 Prediction and Quantization in DPCM	16
2.4 Decorrelating Transformation and Bit Allocation	20
2.5 Adaptive Schemes	23
3 THE BAYESIAN ADAPTIVE DPCM APPROACH	28
3.1 Introduction	28
3.2 Model and Quantization	30
3.3 Limiting Distribution of the Prediction Residuals	34
3.4 Minimum MSE Bit Allocation	45
3.5 Optimal Bit Allocation among Parameters	50
3.6 Simulation and Discussion	61
4 BAYESIAN ADAPTIVE DPCM FOR THE TWO DIMENSIONAL AR(1,1) MODEL	80

4.1	Introduction	80
4.2	Two Dimensional AR(1,1) Model	81
4.3	Two Dimensional Bayesian Adaptive DPCM	87
4.4	Simulation	96
REFERENCES		102
BIOGRAPHICAL SKETCH		106

LIST OF TABLES

3.1	Mean square quantization errors from asymptotic approximation and simulation. $\phi \sim U(0, 1 - \delta)$, $\delta = 10^{-4}$	67
3.2	Mean square quantization errors from asymptotic approximation and simulation. $\pi(\phi) = (1 - \delta)^{-2} 2\phi$, $\delta = 10^{-4}$	72
3.3	Optimal bit number (quantization level) for the parameter vs length of the time series.	77
3.4	Comparison of the mean square quantization errors from different schemes where the series consists several blocks with length $l \sim \text{Poisson}(\lambda)$, and parameter $\phi \sim U(0, 1 - 10^\delta)$, $\delta = 10^{-4}$	78
3.5	Comparison of the mean square quantization errors where the series was generated by using initial parameter $\phi \sim U(0, 1 - \delta)$, $\delta = 10^{-4}$, and a random error $\Delta\phi \sim U(-0.1, 0.1)$ was added after each observation.	79
4.1	Mean square quantization errors from asymptotic approximation and simulation. $\phi_1, \phi_2 \sim U(0, 1 - \delta)$, $\delta = 10^{-2}$	98

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

BAYESIAN APPROACH TO THE ADAPTIVE DPCM FOR THE $AR(P)$ AND
TWO DIMENSIONAL $AR(1,1)$ MODEL

By

FANGSHI SUN

May 1992

Chairman: Mark C. K. Yang
Major Department: Statistics

An analog signal needs to be quantized and encoded into digital form before it is transmitted via a digital channel. A frequently used scheme is the differential pulse-code modulation (DPCM) which predicts the current value by the previous data, and consequently only prediction residuals instead of the original data set need to be transmitted. Usually, DPCM reduces signal redundancy by using an $AR(p)$ time series model to describe the data and to perform the prediction. Due to the nonstationary nature of signals, the parameters in the $AR(p)$ model need to be adjusted from time to time. Several adaptive DPCM schemes have been created to deal with this situation. A different approach motivated by Bayesian analysis is proposed; the parameters are treated as random variables and a prior distribution for the parameters is assumed to be known. For fixed channel capacity, the optimal design of the bit allocation for transmitting the parameters and the prediction residuals is obtained when mean squared quantization error is used as the criterion of the loss function.

CHAPTER 1

INTRODUCTION

1.1 Quantization of Continuous Signals

With the continuing growth of modern communications technology, demand for efficient signal transmission and storage increases rapidly. These practical needs lead to a significant development in theoretical and technical research in the related fields. In recent years, more and more attention has been concentrated on improving the pace and efficiency in signal processing and a lot of research has been conducted in this area. How successfully a communication system works depends both on its "hardware" and its "software." Data compression, concerned with minimizing the number of information carrying units used to represent the signals, is just one of the "software" playing an important role in the operation of a communication system.

Signals we are dealing with often appear in the analog form and they cannot be directly input into digital devices. Suppose x is a single sample from a continuous population with probability density function $f(x)$ defined on a domain R . Since x can take an infinite number of values within a continuous range, it is impossible to transmit the original observation directly via a digital channel due to the limitation of the channel capacity. The data have to be compressed by a quantization procedure in order to pass a finite capacity digital channel. The general practice is to sample the continuous signal at equally spaced intervals of time, and then to digitize each observation to one of several levels predetermined by certain rules. Signal quantization

provides a digital representation for the analog signals. Much of the reason for this use of digital samples of the analog signal is that they can be stored, manipulated and transmitted more reliably than can the analog signal [1]. Nowadays, more and more digital facilities, such as digital computers and digital channels, are used in signal processing and have provided a powerful means for transmitting and storing digital signals in a flexible and reliable manner.

Binary electrical devices are widely used in modern communication. The binary system represents information by only two states which are usually named as "0" and "1". One basic information carrying unit in the binary system, which is called a "bit", provides two stable states which can be used to represent one digit in the binary system. So with m bits available, we can represent 2^m binary numbers.

An analog signal can be considered as a realization from a continuous-time stochastic process with continuous state space. The determination of the optimal quantized values for each observation depends on its statistical behavior.

In mathematically modeling the quantization process, we define a function of the variable x as

$$\hat{x} = q_k \quad \text{if } x \in Q_k \quad (1.1)$$

where $k = 1, 2, \dots, v$, $\{q_k\}$ is a predetermined real sequence and $\{Q_k\}$ is a partition of the range R . The function $\hat{x}$ defined by (1.1) is called a quantizer of x and v is called the quantization level which is determined by the number of bits used to quantize x .

Since it is the quantized values that are actually encoded and transmitted through the channel, it is impossible to restore the original signal exactly. The difference $e = x - \hat{x}$ is called quantization error. The distortion of a quantizer is defined as the expected value of some function of the quantization error, i.e.,

$$D = E[g(x_i - \hat{x}_i)].$$

Among a variety of distortion functions, the mean squared error, with $g(x) = x^2$, is the most frequently used criterion to evaluate the performance of a quantization scheme.

1.2 Classification of the Quantization Schemes

There has been extensive research directed toward signal quantization, and as a result, a large variety of quantization schemes have been developed. According to the ways of handling the quantizer input, these schemes can be classified into three basically different categories [2, 3].

The first category is known as pulse-code modulation (PCM) [4, 5, 6]. In PCM, the sample from the original signal is quantized directly without any previous manipulation. The output of the quantizer is an amplitude discrete sample which will be transmitted through the channel. The advantages of PCM are its simplicity and easy realization. But there is often some kind of redundancy in the signal which is carried on in the quantization and transmission processes if no measure is taken to remove it. So PCM schemes will produce larger quantization error and cost more channel capacity compared with some advanced schemes [7, 8]. Despite the disadvantage, PCM does lay down a basis for all other kinds of schemes.

The second category is called differential pulse-code modulation (DPCM) [9, 10, 11]. DPCM uses the sample intercorrelation to reduce the redundancy carried by the original source signals. In the DPCM system, the current value is predicted by the previous observations, so the sample $\{x_i\}$ can be represented by a model which consists of some parameters plus a residual series $\{\epsilon_i\}$. Instead of encoding the observations themselves, the prediction residuals are encoded and then transmitted via the channel. Since the key function is played by the predictor, DPCM is also called predictive quantization. Due to the reduction of the redundancy by DPCM,

the signal can be transmitted more efficiently than PCM. DPCM can be viewed as a scheme that adopts good features from regular PCM and so is superior to it in performance.

The third category contains all kinds of transform quantization schemes [12, 13]. In these schemes, the sample is divided into data blocks and a transformation is performed in each block. The redundancy is reduced either by an information preserving transformation to yield a new block containing a minimum number of observations [14], or by a decorrelating transformation to make the new block uncorrelated [15]. As they treat a group of data as a whole, the schemes in this category are also called block quantization or vector quantization.

It should be pointed out that some other quantization schemes also exist. They may use a combination of the schemes from the above three categories.

1.3 The Bayesian Approach

In practice, the signals we deal with are often nonstationary. The conventional time-invariant schemes may lose their efficiency in handling nonstationary signals if no adjustment is made. A special but very common type of nonstationarity is caused by the change of the signal features which decide the statistical pattern of the signal. This type of nonstationarity can be described by a model governed by a set of varying parameters and is considered in this paper.

Many kinds of adaptive quantization schemes have been created targeting different adaptation objects. In PCM, the size of the quantization intervals are adapted to match the signal variance. The idea is to modify the interval sizes by a factor depending on the statistical nature of the previous samples and use the modified intervals to quantize the next observation [16, 17]. In DPCM, the parameters are also adapted periodically in addition to modify the quantizer [18, 19, 20]. Once the parameters

are updated by the current observations, the new parameters will be used to perform further predictions. In transform quantization, the adaptation is often made either to the quantizer to fit the statistical nature of the local block or to the design of the bit allocation to improve the quantization efficiency.

In this dissertation, a new approach to quantize nonstationary signals is conducted by combining Bayesian principle with the techniques used in PCM, DPCM and transform quantization. We call it the Bayesian approach. The sample is assumed to come from an autoregressive time series with varying parameters. To overcome the difficulty brought by the nonstationarity, we divide the whole process into blocks with the same size, and stationarity is assumed in each block; i.e., each block forms an autoregressive process with fixed parameters. This is a reasonable assumption if the sampling frequency is large and the block is relatively small.

Inside each block, a modified nonadaptive DPCM procedure is applied based on the fixed parameter model. In a conventional DPCM system, the parameters are first estimated on the transmitter side and the estimate is used to predict the future values of the sample. The difference between the observations and their predicted values are quantized and transmitted to the receiver. Since no bookkeeping information about the parameters is passed over, the parameters need to be estimated again on the receiver side in order to recover the original signal. Though the two estimations are made based on the same formula, the second estimation may contain bias because the quantized data are used on the receiver side. More seriously, the biased estimator may cause further prediction error which may accumulate. It means that any distortion at the receiver may be perpetuated as an error in all future values of the reconstructed signal. In our approach, attempt is made to avoid the error accumulation in the signal reconstruction process by transmitting the prediction residuals as well as the quantized parameters. Motivated by the principle of the Bayesian analysis [21], we treat the varying parameters as random variables and assigned them a joint prior

density function based on the previous knowledge about the process. Since only the quantized values of the parameters are available for the signal reconstruction, it is advantageous to use the quantized parameters for prediction on the transmitter side. The whole procedure can be stated as follows. We quantize the parameters and then use the quantized parameter to perform prediction. The quantized values of the parameters are transmitted to the receiver as well as the quantized residuals. The predictor on the receiver side will work on these values to recover the original signal. If the transmission channel is errorless, the distortion of the recovered signal is just the quantization errors of the residuals which are generally quite small.

In the new scheme, the quantization of the residuals is closely related to that of the parameters. High level-quantized parameters will reduce the variance of the residuals and so increase the accuracy of the prediction. On the other hand, high level parameter quantization will cost more channel capacity and leave fewer bits available for quantizing the residuals. To find the optimal bit allocation, we need to examine how seriously the reconstruction is affected by the quantization of the parameters. Since the blocks are treated independently in our scheme, there is at least one nonpredictable observation in each block that has no previous information. Thus, it is unwise to distribute the bits equally among all the observations without considering the difference in predictability. Therefore, the bits allocation among the parameters, predictable and nonpredictable observations needs to be determined in order to obtain the minimum distortion.

As pointed out above, the quantized parameters are used in our prediction procedure, so the residuals are no longer uncorrelated. Under this circumstance, it seems difficult to obtain the exact theoretical result. In this dissertation, an effort is made to get a close theoretical approximation for the residuals based on their limiting behavior as the quantization levels tend to infinity. Simulation is performed to verify

the closeness of this approximation when the quantization levels are small. Comparison is made with the conventional DPCM scheme and another adaptive DPCM scheme.

CHAPTER 2

QUANTIZATION SCHEMES

2.1 Introduction

As mentioned in the previous chapter, the variety of the quantization schemes can be generally classified into three categories, namely, PCM, DPCM and transform quantization according how they dispose of signals before inputting them to the quantizer.

PCM uses the original sample as the quantization input and obtains a digital representation from the quantization output. The research in this field was concentrated on finding the optimal quantizer to minimize certain type of distortion measure. In 2.2, an optimal quantizer, in the sense of minimizing the mean squared quantization error, is introduced and its properties are discussed. Except variables with uniform distribution, the functional form of the quantizer is often too complicated to be expressed analytically. An iterative logarithm is suggested to obtain the numerical solution. It is also desirable to express the quantization error as a function of the quantization levels so that further analysis can be conducted. An approximation based on large quantization level is derived in this section.

DPCM makes an effort to improve the prediction as well as to optimize the quantization. The two main procedures of DPCM, namely prediction and quantization, are briefly discussed in 2.3. In DPCM, a statistical model is needed to perform the prediction. The widely used autoregressive model is described and the estimation

and prediction with this model are considered. After the prediction residuals are obtained, the results obtained from PCM can be applied in the quantization procedure in DPCM.

Transform quantization is an alternative to DPCM to improve the quantization input. Two main concerns in this scheme are the selection of transform matrix and the allocation of the bits. Both aspects are considered in 2.4. Several transform methods are given and their optimality and complexity are discussed. The bit allocation problem arises as the components of the transformed sample may not have identical variance. An optimal design for the bit allocation is presented in this section.

All schemes mentioned above can be made adaptive to fit nonstationary samples. In DPCM, either the predictor or the quantizer can be updated by using the newly observed data. In transform quantization, the adaptation is achieved by the adjustment of the transform matrix or by changing the bit assignments for different blocks. Adaptive quantization is the topic discussed in 2.5.

The content reviewed in this chapter forms the theoretical basis for our new approach.

2.2 Minimum Mean Square Error Quantization

PCM is a simple quantization scheme in digital transmission but it provides some necessary means and techniques to other more comprehensive schemes. In their early paper in 1948, Oliver, Pierce and Shannon elaborated the philosophy of PCM and pointed out the differences between what can be achieved with PCM and with some other previous applied systems [4].

As mentioned previously, PCM is a scheme which gives a discrete digital representation to the analog signals. Since the original signal cannot be fully recovered from its quantized values [22], great effort has been made to find an optimal scheme

to minimize the distortion. In our approach, the mean square error (MSE) is used as the criterion to compare different schemes.

Minimum mean squared error quantizer

In the early sixties, Lloyd [23] and Max [5] individually solved the problem of optimal quantization for operation on a single sample of a random process with some specified probability density.

Let X be a random variable with density function $f(x)$ defined on the domain R and x be an observation. To obtain the optimal quantization scheme which minimizes the MSE for a given quantization level v , one just needs to find a partition $\{Q_k\}$ and the corresponding quantization values $\{q_k\}$ such that

$$\begin{aligned} D &= E[(X - \hat{X})^2] \\ &= \sum_{k=1}^v \int_{Q_k} (x - q_k)^2 f(x) dx \end{aligned} \quad (2.1)$$

is minimized.

We now consider the minimization problem in two steps. First, we assume that the partition $\{Q_k\}$ has already been chosen. A well known result in statistics tells us that

$$\int_{Q_k} (x - q_k)^2 f(x) dx$$

attains its minimum if and only if $q_k = E(X|X \in Q_k)$. This is to say, the best choice of $\{q_k\}$ is just the center of the mass of X in $\{Q_k\}$ with density $f(x)$.

Next consider the problem of finding the best $\{Q_k\}$ given that the values $\{q_k\}$ are fixed. Without loss of generality, assume $q_1 < q_2 < \dots < q_v$. Since $(x - q_k)^2$ has positive weight $f(x)$ in the expression of the mean squared quantization error, it is

natural that any value closer to q_k than to any other quantization value should be assigned to Q_k . Ignoring a subset with zero probability, Q_k should be chosen as

$$Q_k = \{x : (x - q_k)^2 < (x - q_j)^2, \text{ for all } j \neq k\}.$$

The boundary point can be assigned to either of the two sets it separates without affecting the mean square quantization error.

It is straightforward that $\{Q_k\}$, the partition of R , should be intervals whose endpoints, say $c_k, k = 0, 1, \dots, v$, bisect the segments between successive $\{q_k\}$, except c_0 and c_v which can be any values such that $-\infty < c_0 \leq \inf(x : x \in R)$ and $\sup(x : x \in R) \leq c_v < \infty$. So the best partition $\{Q_k\}$ should be a selection of intervals with endpoints $\{c_k\}$. By convention, we assign the boundary point to the interval on its left, i.e.

$$\{Q_k\} = \{(c_{k-1}, c_k]\}.$$

From the above discussion, we have obtained the following necessary condition for $\{Q_k\}$ and $\{q_k\}$ to be optimal.

$$\begin{cases} \int_{c_{k-1}}^{c_k} (x - q_k) f(x) dx = 0 \\ c_k = (q_k + q_{k+1})/2 \end{cases} \quad k = 1, 2, \dots, v. \quad (2.2)$$

The above equations can also be obtained by differentiating the total mean square quantization error D defined in (2.1) with respect to $\{q_i\}$ and $\{c_i\}$, then setting the derivatives to zero.

The sufficient condition is that the Hessian matrix of $E[(X - \hat{X})^2]$ is positive definite.

The L-M quantizer has the following properties which can be easily proved [24].

- (1) $E(\hat{X}) = E(X)$.
- (2) $E(X\hat{X}) = E(\hat{X}^2)$.
- (3) $Cov(X, X - \hat{X}) \geq 0$.

The quantized values and the end points of the quantization intervals for the standard normal distribution is given by Max [5], and they were also obtained for the gamma and Laplacian distribution [25].

A numerical algorithm

Because of the complicated functional relationships which are likely to be induced by $f(x)$, the analytic solution of the simultaneous equations (2.2) may not be easily obtained. Lloyd and Max introduced a numerical method to solve this problem [5, 23].

Choose $-\infty = c_0^{(1)} < c_1^{(1)} < \dots < c_{v-1}^{(1)} < c_v^{(1)} = \infty$ arbitrarily as the initial values and then let $\{q_i^{(1)}, i = 1, 2, \dots, v\}$ be the center of the mass of x in $(c_{i-1}^{(1)}, c_i^{(1)})$. By the above discussion, the optimal choice of $\{c_i\}$ should be the center points of the intervals $\{(q_i, q_{i+1})\}$ which in general may not be satisfied by the arbitrary choice of $\{c_i^{(1)}\}$. So the second trial will choose that

$$c_i^{(2)} = \frac{(q_i^{(1)} + q_{i+1}^{(1)})}{2}, \quad \text{for } i = 1, 2, \dots, v-1$$

and correspondingly

$$q_i^{(2)} = E[X | c_{i-1}^{(2)} < X < c_i^{(2)}].$$

Note that the values for c_0 and c_v remain unchanged throughout the whole procedure.

The values of $\{c_i^{(2)}\}$ and $\{q_i^{(2)}\}$ are obtained as the result of imposing the center point condition and center of mass condition again. It is obvious that the resulting mean squared error after the second step of the iteration will become smaller, i.e.,

$$\text{MSE}^{(2)} < \text{MSE}^{(1)}.$$

Continuing this process, we get a sequence of values for the MSE. After the m th step, we have

$$\text{MSE}^{(m)} < \text{MSE}^{(m-1)} < \dots < \text{MSE}^{(1)}.$$

Since it is a nonnegative and decreasing sequence, the limit of $\{\text{MSE}^{(m)}\}$ exists as m tends to infinity [26], and consequently, the limits of $\{c_i^{(m)}\}$ and $\{q_i^{(m)}\}$ should possess the optimal property.

Some approximations

Though numerical method is available, it is still desirable to get a close form expression for $\{c_k\}$ and $\{q_k\}$. An analytical approximations for $\{c_i\}$ and $\{q_i\}$ was obtained by Roe [27]. When X has a standard normal distribution,

$$c_k = \sqrt{6} \operatorname{erf}^{-1} \left(\frac{2k - v}{v + 0.8532} \right), \quad (2.3)$$

where erf^{-1} is the inverse error function.

Once the endpoints of the quantization intervals are obtained, the approximated values of $\{q_i\}$ can be obtained from (2.2).

Though the approximation is based on large quantization level, it works quite well even for moderate value of v . For the case of $v = 6$, the individual values of c_k

calculated from (2.3) are within 3 percent of the correct values. For $v = 36$, they are within 0.5 percent and the resulting quantization error is undistinguishable from the true error.

It is helpful to obtain the relationship between the MSE and the quantization level for studying the error rate. Max showed by a graphing argument that

$$E[(X - \hat{X})^2] \simeq k(v)v^{-2}$$

approximately, where $k(v)$ is some function of v .

An approximation for the quantization error, when v is large enough, was given by Panter and Dite [28]. If the quantization level v is large, $f(x) \simeq f(q_i)$ for $x \in Q_i, i = 1, 2, \dots, v$ and q_i is almost the center point of the interval $(c_{i-1}, c_i]$. Let $\Delta_i = q_i - c_{i-1} = c_i - q_i$, then

$$\begin{aligned} E[(X - \hat{X})^2] &= \sum_{i=1}^v \int_{c_{i-1}}^{c_i} (x - q_i)^2 f(x) dx \\ &\simeq \sum_{i=1}^v \frac{1}{3} f(q_i) [(c_i - q_i)^3 + (q_i - c_{i-1})^3] \\ &= \frac{2}{3} \sum_{i=1}^v f(q_i) \Delta_i^3 \\ &= \frac{2}{3} \sum_{i=1}^v u_i^3 \end{aligned} \tag{2.4}$$

where $u_i = f^{\frac{1}{3}}(q_i) \Delta_i$. Hence,

$$\sum_{i=1}^v u_i \simeq \frac{1}{2} \int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx = C, \quad (2.5)$$

where C is some constant.

This problem is now reduced to minimizing the sum of cubes subject to the condition that the sum of the variables is a constant. From Lagrange's method of undetermined multipliers, it follows that, subject to (2.5), the expression (2.4) is minimized when

$$u_1 = u_2 = \dots = u_v = \frac{C}{v}.$$

Therefore,

$$E[(X - \hat{X})^2] \simeq \frac{2}{3} \sum_{i=1}^v \frac{C^3}{v^3} \simeq \frac{1}{12v^2} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^3. \quad (2.6)$$

This formula gives an explicit analytic approximation for the MSE as a function of the quantization level v . Lloyd obtained the same result from a different approach [23]. Lloyd also proved that the difference between the approximation and the MSE has a higher order of magnitude than v^{-2} , i.e.

$$E[(X - \hat{X})^2] - \frac{1}{12v^2} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^3 = o(v^{-2}). \quad (2.7)$$

It is trivial that a higher quantization level will produce a smaller mean squared error. When the quantization level v is given, the lower bound for the mean squared quantization error for any quantizer is given by [6]

$$E[(X - \hat{X})^2] \geq \frac{E(X^2)}{2^{2v}}.$$

The above lower bound can be used to evaluate the performance of any quantizer.

2.3 Prediction and Quantization in DPCM

The operation of a typical DPCM system can be described as follows. A predictor on the transmitter side works on the previous data and produces a prediction for the present value. The difference between the true value and the prediction is then quantized and transmitted to the receiver. On the receiver side, another predictor works synchronously with the one on the transmitter side. The prediction generated by this predictor is added to the received difference to reconstruct the original signal. The variance of the differences input to the quantizer is generally smaller than that of the original signal because the redundancy concealed in the sample has been removed. Therefore, it provides a possible way to save the bandwidth at an acceptable distortion rate or obtain smaller distortion at fixed channel capacity. DPCM is more powerful and more efficient to quantize and transmit highly correlated signals than the PCM scheme since we have taken advantage of the knowledge about intersample correlation in the prediction procedure. This is the key point making DPCM superior to PCM.

DPCM was first revealed by Cutler of the Bell Telephone Laboratory in his original patent in 1952 [29]. Since then, much research has been conducted to improve and consummate this scheme. Elias discussed this scheme in detail and suggested some criterion for the optimal predictor [10]. Goldstein and Liu found the joint distribution of the quantization error and the step-size of the uniform quantizer adopted in the system based on the assumption that the sample forms a first order Markov sequence [30]. Arnstein formulated the quantization error by a difference equation and suggested an iterative algorithm to calculate the numerical result [31].

In DPCM, the sample needs to be modeled in order to perform predictions. The model used by DPCM is an autoregressive process, i.e., a time-variate process with

each observation drawn from some probability distribution conditioned on the previous observations. When the current value is a linear combination of the previous p observations plus some white noise, it is called a p th order autoregressive time series (AR(p)) [32, 33]. Practical experience has shown that the AR(p) model fits a large variety of data very well. Much of the literature about coding and quantization is concentrated on this model [18, 34, 35].

Consider a stationary autoregressive sequence $\{x_i, \pm i = 0, 1, 2, \dots\}$ with common mean zero and variance σ_x^2 . An AR(p) model is defined by

$$x_i = \phi_1 x_{i-1} + \phi_2 x_{i-2} + \dots + \phi_p x_{i-p} + \epsilon_i, \quad (2.8)$$

where $\{\epsilon_i\}$ is the white noise sequence.

In practice, $\{x_i\}$ should be able to be written as a function of the previous noise sequence in order to have a realistic physical interpretation [33]. For a process to possess this property, the parameters $\{\phi_j\}$ should satisfy the condition that the complex equation

$$1 - (\phi_1 z + \phi_2 z^2 + \dots + \phi_p z^p) = 0,$$

has no solution z on or inside the unit circle. The processes satisfying this condition are called causal.

The lag h autocovariance and correlation functions are defined by

$$\gamma_h = Cov(X_i, X_{i+h})$$

and

$$\rho_h = \frac{Cov(X_i, X_{i+h})}{Var(X_i)} = \frac{\gamma_h}{\gamma_0}$$

respectively. The most frequently used estimators for γ_h and ρ_h are the sample covariance and sample correlation given by

$$\hat{\gamma}_h = \frac{1}{n} \sum_{i=1}^{n-h} (X_i - \bar{X})(X_{i+h} - \bar{X})$$

and

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0}$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ is the sample mean.

An estimate for the parameters can be obtained by substituting the sample correlation for the true correlation in the famous Yule-Walker equation [33]

$$\hat{\Phi} = \hat{P}^{-1} \hat{\rho},$$

where

$$\hat{\Phi} = (\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p)',$$

$$\hat{\rho} = (\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_p)'$$

and

$$\hat{P} = \begin{pmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-2} \\ \hat{\rho}_2 & \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{p-3} \\ \vdots & & & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \cdots & 1 \end{pmatrix}.$$

We now consider the prediction problem based on the model we have established. In view of the demand in DPCM, we are interested only in one step ahead prediction. That is, we want to predict the value of x_{n+1} based on our knowledge about $\{x_n, x_{n-1}, \dots\}$. It is well known that the minimum MSE predictor [33] is

$$\hat{X}_{n+1} = E[X_{n+1} | X_n, X_{n-1}, \dots].$$

Recall the Markov property that the distribution of X_{n+1} depends only on the p preceding observations. So the MSE predictor for the AR(p) model is

$$\begin{aligned}\hat{X}_{n+1} &= E[X_{n+1}|X_n, X_{n-1}, \dots, X_{n-p+1}] \\ &= \phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_p X_{n-p+1}.\end{aligned}\quad (2.9)$$

The prediction residual is given by

$$\begin{aligned}X_{n+1} - \hat{X}_{n+1} \\ &= X_{n+1} - (\phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_p X_{n-p+1}) \\ &= \epsilon_{n+1}.\end{aligned}\quad (2.10)$$

which is the noise term at lag $n + 1$. This means that all the redundancy has been removed and only the uncontrollable random factor is left.

Note that

$$\begin{aligned}Var(X_{n+1}) &= Var(\phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_p X_{n-p+1} + \epsilon_{n+1}) \\ &= Var(\phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_p X_{n-p+1}) + \sigma_\epsilon^2 \\ &\geq \sigma_\epsilon^2.\end{aligned}$$

The difference between $Var(X_{n+1})$ and σ_ϵ^2 is generally significant, specially if the process has highly positive correlations among the successive observations. A precise relation between the two variances is

$$Var(X_i) = \frac{\sigma_\epsilon^2}{(1 - \phi_1 \rho_1 - \phi_2 \rho_2 - \dots - \phi_p \rho_p)}.$$

This tells us the improvement of the quantizer input for the DPCM system compared to that for the PCM system.

Since the prediction residuals are statistically uncorrelated, nothing can be done for further improvement in modeling. What is needed now is to choose a quantizer working on the prediction errors. From the statistical point of view, the optimal choice is the L-M quantizer as mentioned previously. When a large number of bits are available, an alternative choice, without much loss in distortion, is the uniform quantizer with equal size quantization intervals except the first and the last one. Other quantizers are also used in DPCM [36].

Some other quantization schemes were created under criteria other than the MSE. O'Neal suggested a scheme aimed at maximizing the sample entropy and showed that the entropy quantization can increase the signal-to-noise ratios [11]. Some more work was done for getting better subjective signal quality [37, 38].

Research has been conducted to reveal the nature of the quantization errors coming from different types of quantizers. Due to the nonlinear nature of the quantizers, the exact analysis of the quantization error seems hard to carry on except for in some special cases. Most works reported in the literature are using some kind of approximation based upon the limiting behavior.

2.4 Decorrelating Transformation and Bit Allocation

An alternative approach to handle the redundancy in the sample is the transform quantization [12, 39, 40]. Because it deals with the quantizations of a group of observations at the same time rather than quantizing them individually, the optimal block quantizer refers to the one minimizing the total mean squared error of the whole block with fixed total number of bits. The transformation removes the redundancy by

transforming one block to another either with uncorrelated components [15] or with fewer number of components for quantization [14]. Then, a corresponding inverse transformation is performed on the receiver side to restore the original block.

Linear transformation is most frequently used for decorrelating the sample [15, 41]. Assume $\mathbf{X} = (x_1, x_2, \dots, x_n)'$ is a random vector with mean vector $\mathbf{0}$ and covariance matrix Σ . The vector $\mathbf{X}$ is linearly transformed by a $n \times n$ matrix A to produce another vector $\mathbf{Y}$ with uncorrelated components which are quantized with a chosen quantizer and then transmitted via the channel. Another linear transformation on the receiver side by an $n \times n$ matrix B is performed to yield the vector $\hat{\mathbf{X}}$ as the reconstructed version of $\mathbf{X}$. The problem is to choose the matrices A and B such that

$$D = E \sum_{i=1}^n [(x_i - \hat{x}_i)^2]$$

attains its minimum.

Huang and Schultheiss studied this problem and obtained the optimal choice of the matrices A and B [15]. Suppose A is chosen, the optimal B is simply the inverse matrix of A , i.e., $B = A^{-1}$. The best choice of the transform matrix A satisfies

$$A \Sigma A' = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

where diag denotes the diagonal matrix and $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of the covariance matrix Σ in descending order. It is easily seen that the rows of A are eigenvectors of Σ . This transformation by A and B is known as Karhunen-Loève (K-L) transform.

Though K-L transform has the optimum property, it presents some problems in practical application [2, 3]. First, the covariance matrix is often unknown. Using an estimate from the previous block may lead to a loss in optimality if the sample is

nonstationary. Second, the covariance matrix sometimes appears to be near singular, so its eigenvectors are not uniquely determined. Furthermore, the computation is difficult and there is no fast algorithm associated with it.

Discrete cosine transform (DCT), which is a member of a large family called sinusoidal transform, is frequently used in practice [42]. DCT has a performance close to the K-L transform but is much easier in implementation.

The transform matrix A in DCT is defined by

$$a_{ij} = \frac{2c_i}{\sqrt{n}} \cos\left[\frac{(2j+1)i\pi}{2n}\right]$$

where a_{ij} is the entry in the i th row and j th column and

$$c_i = \begin{cases} 1 & \text{for } i = 1 \\ \frac{1}{\sqrt{2}} & \text{for } i = 2, 3, \dots, n. \end{cases}$$

Other members in the sinusoidal family such as discrete Fourier transform and Hadamard transform have equivalent performance as DCT [43].

Once A and B are defined, another problem is to determine the bit allocation among the components in the block. Huang and Schultheiss studied this problem and obtained an optimal solution [15]. Assume b bits are available for quantizing the whole block and b_i denotes the number of bits assigned to the i th component. The mean squared error can be written as

$$E\Sigma_{i=1}^n [(x_i - \hat{x}_i)^2] \simeq \sum_{i=1}^n K 2^{-2b_i \sigma_i^2}, \quad (2.11)$$

Set the derivative of the above expression to zero subject to $\Sigma_{i=1}^n b_i = b$. The optimal bit allocation design is given by

$$b_i = \frac{b}{n} + \frac{1}{2} \log_2 \frac{\sigma_i^2}{[\text{Det}(\Sigma)]^{\frac{1}{n}}} \quad (2.12)$$

where Det denotes the determinant of a square matrix.

The solution for (2.12) sometimes contains unrealistic negative values. An adjustment was made by Segall to ensure a nonnegative solution [44].

The constant K in (2.11) is greater than 1. Substituting (2.12) into (2.11), we can get a lower bound for the average mean square error by using Huang and Schultheiss' scheme,

$$\frac{1}{n} E \sum_{i=1}^n [(x_i - \hat{x}_i)^2] \geq 2^{-\frac{2K}{n}} [\text{Det}(\Sigma)]^{\frac{1}{n}}.$$

Another approach to block quantization was conducted by quantizing the whole block as a vector and the quantization is performed within the corresponding vector space. Fischer and Dicharry [45] suggested a locally optimal vector quantizer for Gaussian, gamma and Laplacian sources and showed that in low dimensional case (2-6), the improvement in MSE is significant compared to scalar quantizer for gamma and Laplacian sources. Vector quantization usually requires complicated computation and some research was done to reduce the complexity [46, 47].

2.5 Adaptive Schemes

The conventional DPCM and transform schemes described above are used to quantize stationary signals. Some adaptation needs to be made when dealing with nonstationary signals in order to keep the efficiency of the original scheme [48]. In each category of quantization schemes mentioned above, adaptation can be made aiming at different adapting objects.

A DPCM system consists of two key components, namely, the predictor and the quantizer, both of which can be made adaptive. Predictor adaptation is done by changing the prediction parameters in order to fit the statistical behavior of the local sample [19, 20]. In these schemes, the parameters are adjusted every period of certain length and then, the new parameters are used in the prediction for the next period of time. The quantizer is often adapted by altering the sizes of the quantization intervals which will be enlarged or reduced by a certain measurement in accordance to the variation of the previous sample [16, 49].

The adaptive quantizer has attracted more attention than the adaptive predictor because of its easier realization. Most of the previous coding literature has been limited to linear predictors with fixed parameters, and based on this assumption, the emphasis has been put on the improvement of the performance of different types of quantizers. Cummiskey, Jayant and Flanagan [48] and Jayant [16] discussed a uniform adaptive quantizer which adapts the interval size according to which slot of the quantizer the previous sample falls in. The new interval size is determined by increasing or decreasing a multiple of the old interval size by a time-invariant function. Let Δ_i be the interval size for quantizing observation x_i , then,

$$\hat{x}_i = P_i \frac{\Delta_i}{2} \quad \pm P_i = 1, 3, \dots, 2^b - 1$$

if a b -bits quantizer is used, and

$$\Delta_i = \Delta_{i-1} M(|P_{i-1}|).$$

The adaptive multiple M is determined by the interval that the previous quantized value belongs to. $M < 1$ is assigned for small value of $|P_{i-1}|$, which means x_{i-1} falls into an interval located near the center of its range. On the other hand, if x_{i-1} falls into an interval near the edge of its range, i.e., the value of $|P_{i-1}|$ is large, we will take $M > 1$. The desire values of M were found by computer simulation.

Note that the adaptation scheme is known from the information on the receiver end, so no adaptive parameter needs to be passed.

Another adaptive quantizer was created by Noll in which the adaptation is controlled by the adjusted estimate of the sample variance [50].

In practice, to control the prediction error is sometimes more efficient and more important. Because the prediction error accumulates during the whole transmission process, any present error will affect all future predictions. Thus, the handling of the parameter variation might affect the prediction error greatly. Several predictor adaptive DPCM schemes have been created to deal with this case.

Cummiskey suggested an adaptive predictor obtained by steepest descent gradient search which can be illustrated as follows [18]. Suppose $x_1, x_2, \dots, x_n$ are a sample from an $AR(p)$ process. The optimal predictor for x_i is $\sum_{j=1}^p \phi_j x_{i-j}$, and the corresponding squared prediction residual is $e_i^2 = (x_i - \sum_{j=1}^p \phi_j x_{i-j})^2$. The gradient of e_i^2 with respect to the parameters $\{\phi_j\}$ is given by

$$\text{Grad}(e_i^2) = \begin{pmatrix} \frac{\partial e_i^2}{\partial \phi_1} \\ \frac{\partial e_i^2}{\partial \phi_2} \\ \vdots \\ \frac{\partial e_i^2}{\partial \phi_p} \end{pmatrix} = - \begin{pmatrix} 2e_i x_{i-1} \\ 2e_i x_{i-2} \\ \vdots \\ 2e_i x_{i-p} \end{pmatrix}.$$

By definition, the direction of the gradient is where e_i^2 will attain maximum increment. So $\{\phi_j\}$ should be adapted in the direction opposite to the gradient. The increments, after normalization, become

$$\Delta \phi_j = \frac{K e_i x_{i-j}}{\sum_{i=1}^p x_{i-j}^2} \quad (2.13)$$

where K is the adapting rate.

Cummiskey also suggested that if the bit rate is not so small ($b > 2$), the transmission of the prediction coefficients can be avoided by determining these coefficients from the reconstructed signal on the receiver side, and the cost for doing this is rather small. Atal and Schroeder described a scheme in which the prediction parameters are periodically readjusted and suggested a bit rate for the transmission of the residuals together with the parameters. But no attempt was made to optimize the quantization of the parameters [19].

In transform quantization, adaptation can be made either to the transformation matrix or to the bit allocation design. The adaptive transform matrix can remove the local redundancy and feed the quantizer with better input. The adaptive bit allocation can improve the quantization efficiency with fixed bit rate.

An adaptive transform method, aiming at adapting the transform matrix A , was proposed by Tasto and Wintz [49, 51]. Assume $\mathbf{X} = (x_1, x_2, \dots, x_n)'$ is a random vector with

$$\begin{aligned} \mathbf{U}_{\mathbf{X}} &= E(\mathbf{X}) \quad \text{and} \\ C_{\mathbf{X}} &= E[(\mathbf{X} - \mathbf{U}_{\mathbf{X}})(\mathbf{X} - \mathbf{U}_{\mathbf{X}})']. \end{aligned}$$

The transform matrix A needs to be adjusted according to the variation of the covariance matrix $C_{\mathbf{X}}$. The adaptation is made by partitioning the sample space Ω into L disjoint subsets $\{S_i\}$ with corresponding probability $\{P_i\}$ such that the vectors from the same subset are expected to have homogeneous statistical characteristics and will be transformed by the same matrix A_i . Now, the problem is how to find the optimal partition such that the distortion is reduced to minimum. Tasto and Wintz suggested that the subset S_k should contain those $\mathbf{X}$ for which

$$D_k(\mathbf{X}) = \min_{1 \leq i \leq L} D_i(\mathbf{X})$$

where

$$D_i(\mathbf{X}) = \alpha_i + (\mathbf{X} - \mathbf{U}_i)' \mathbf{C}_i^{-1} (\mathbf{X} - \mathbf{U}_i)$$

$$\mathbf{U}_i = E(\mathbf{X} | S_i)$$

$$\mathbf{C}_i = E[(\mathbf{X} - \mathbf{U}_i)(\mathbf{X} - \mathbf{U}_i)' | S_i]$$

and α_i is some constant.

The optimal choices of L and P_i were suggested according to simulation results.

Another adaptive transform quantization scheme focusing on the adaptation of the bit assignment was suggested by Chen and Smith [14]. In this scheme, the image block is classified into several classes according to a measure of block activity. More bits will go to the active blocks than the quiet ones. Within each block, the bits are distributed according to the covariance matrix of the transformed data by using Huang and Schultheiss' scheme.

Cuperman and Gersho [52] proposed an adaptive scheme in which a low-dimensional vector quantizer is used in an adaptive predictive coding scheme. The current input vector is predicted and the residual vector is coded by a vector quantizer.

CHAPTER 3

THE BAYESIAN ADAPTIVE DPCM APPROACH

3.1 Introduction

A new approach, which we called Bayesian adaptive DPCM, is attained in this chapter to deal with nonstationary sequences. We are concerned with the nonstationary p th order autoregressive processes with varying parameters which can be modeled by

$$x_t = \sum_{i=1}^p \phi_{it} x_{t-i} + \epsilon_t. \quad (3.1)$$

The new scheme, considering the quantization for both parameters and the residuals, is developed using the techniques in DPCM and block quantization combined with the Bayesian principle. To overcome the nonstationarity, we divide the time series into small segments and sample is drawn from each of them. If the sampling frequency is high and the segment is small, it is reasonable to assume the sample from this segment is stationary. The signal then is decomposed as a series of stationary subsequences and each of them is governed by a set of constant parameters. The parameters are estimated inside each sample and then used to make prediction. The residuals and the parameters are treated as a block in the quantization and transmission processes. The parameters, whose values vary from sample to sample, are treated as random variables. Motivated by Bayesian analysis, a prior density

function based on the previous experience is assigned to the parameters and used to determine the optimal quantization.

Bit allocation needs to be considered in block quantization since each component in the block makes different contribution to the distortion. In our approach, the components in the block can be divided into three groups. The first group contains all parameters. The quantization of the parameters will affect the performance of the prediction. Using low level quantized parameters increases the variance of the residuals, but, on the other hand, high level in parameter quantization requires more channel capacity and leaves fewer bits available for quantization of the residuals. So they need special attention in the quantization procedure. The second group contains the first p observations in each sample. As each block is treated individually, the first p observations at the beginning of the block are not predictable, and they have to be quantized directly. The third group is formed by all prediction residuals. Since they generally have smaller variance than the components in the second group, it may not be optimal to equally distribute the available bits among them.

In Bayesian analysis, a loss function is needed as the criterion of the performance of the procedure. For our problem, the total mean squared error is naturally taken to play this role. Our research in this paper is trying to obtain the optimal design for bit allocation among the parameters $\{\phi_1, \phi_2, \dots, \phi_p\}$, the first p observations and the other prediction residuals to minimize the total mean squared quantization error under the assumption that each sample forms an stationary $AR(p)$ sequence.

In 3.2, the model and quantization procedure adopted in our new scheme are introduced and some related properties are discussed. In 3.3, an approximation for the variance of prediction residuals is derived under the consideration of the parameter quantization. The limiting behavior of the residuals is closely examined. An alternative proof for the optimal bit allocation for block quantization is given in 3.4. The above results are applied in 3.5 and the optimal bit allocation for the

Bayesian DPCM is obtained. Simulation is performed and the result is compared with other schemes in 3.6.

3.2 Model and Quantization

We have mentioned in Chapter 1 that samples from different segments of the signal are treated individually and the association between segments are not considered. Let $\{x_1, x_2, \dots, x_n\}$ be a sample drawn from an p th order autoregressive process. Since the prediction is based only on a finite number of observations, a truncated version of the $AR(p)$ model is adopted in our scheme, which is defined by

$$\begin{aligned} x_i &\sim N(0, \sigma_x^2), \quad i = 1, 2, \dots, p, \\ x_i &= \Phi' X_{i-1} + \epsilon_i, \quad i = p+1, p+2, \dots, n, \end{aligned} \quad (3.2)$$

where ϵ_i 's are independent and identically distributed (i.i.d.) with normal distribution $N(0, \sigma_\epsilon^2)$. $\Phi' = (\phi_1, \phi_2, \dots, \phi_p)$ and $X_{i-1}' = (x_1, x_2, \dots, x_n)$.

Suppose Φ has known prior density function $\pi(\phi_1, \phi_2, \dots, \phi_p)$ defined on the admissible region by which we mean a set of points with coordinates $(\phi_1, \phi_2, \dots, \phi_p)$ which make the model (3.2) stationary. It is well known that for $p = 1$ and $p = 2$, the admissible regions are

$$\{\phi : |\phi| < 1\}$$

and

$$\{(\phi_1, \phi_2) : \phi_2 + \phi_1 < 1, \phi_2 - \phi_1 < 1, \text{ and } |\phi_2| < 1\}$$

respectively. Furthermore, we assume the parameter vector Φ is independent of the white noise $\{\epsilon_i\}$.

The first p observations, $x_1, x_2, \dots, x_p$, are not predictable due to the lack of previous information. To simplify the notation, we define the first p prediction errors to be equal to the observations and they will be quantized directly. Model (3.2) is used to predict future observations $x_i, i = p+1, p+2, \dots, n$, and the residuals will be quantized and transmitted. By definition, the residuals can be written as

$$e_i = \begin{cases} x_i, & \text{if } i = 1, 2, \dots, p; \\ x_i - \hat{\Phi}'\hat{X}_{i-1}, & \text{if } i = p+1, p+2, \dots, n, \end{cases} \quad (3.3)$$

where $\hat{\Phi}$ and $\hat{X}_{i-1}$ denote the quantized values of Φ and X_{i-1} respectively. The same notation will be used for other random variables throughout this paper unless otherwise stated. To avoid confusion, it should be pointed out that $\hat{X}_i$ is obtained by adding $\hat{e}_i$ to its predicted value rather than being quantized directly. Therefore, $x_i - \hat{x}_i$ is called a reconstruction error instead of a quantization error.

The quantized values of the prediction errors are transmitted via the channel. If the channel is error-free, it is easy to show that the reconstruction error $x_i - \hat{x}_i$ is simply equal to the quantization error $e_i - \hat{e}_i$. For $i \leq p$, this is trivial by definition. For $i > p$,

$$x_i - \hat{x}_i = (x_i - \hat{\Phi}'\hat{X}_{i-1}) + (\hat{x}_i - \hat{\Phi}'\hat{X}_{i-1}) = e_i - \hat{e}_i.$$

The optimality of the L-M quantizer has been shown in the previous chapter and is chosen to perform the quantization in our new scheme. The quantized values and the quantization intervals for the standard normal variable are given by Max [5]. For the general normal variable, the elements of the quantizer can be obtained by the next lemma.

Lemma 3.2.1

Assume X is a random variable with MSE quantizer $\hat{X}$. Let a and b be real constants. Then, the MSE quantizer for $Y = a + bX$ is given by $\hat{Y} = a + b\hat{X}$, and the mean square quantization error is given by

$$E[(Y - \hat{Y})^2] = b^2 E[(X - \hat{X})^2].$$

Proof of Lemma 3.2.1 :

Let $\tilde{Y}$ be any quantizer of Y , we have

$$E[(Y - \tilde{Y})^2] = b^2 E[(X - \frac{\tilde{Y} - a}{b})^2] \geq b^2 E[(X - \hat{X})^2].$$

The equality holds if and only if

$$\frac{\tilde{Y} - a}{b} = \hat{X}.$$

So we get

$$\tilde{Y} = a + b\hat{X}.$$

This lemma allows us to assume the variable x has zero mean and unit variance without loss of generality.

We will use the properties of the L-M quantizer stated in 2.2 in the derivation of the new scheme. Specially property (2) will be used repeatedly in different forms. It is helpful to give all of the presentations of this property.

Lemma 3.2.2

Let X be a random variable and $\hat{X}$ the L-M quantizer. Then, the following statements are equivalent.

1. $E(X\hat{X}) = E(\hat{X}^2)$.
2. $E[(X - \hat{X})\hat{X}] = 0$.
3. $E[(X - \hat{X})X] = E[(X - \hat{X})^2]$.
4. $E[(X - \hat{X})^2] = E(X^2) - E(\hat{X}^2)$.

The proof is easy and so is omitted.

As mentioned in Chapter 2, the L-M scheme is created to quantize a single observation or a set of uncorrelated variables. We need to consider the quantization of the parameters which may not necessarily be uncorrelated. The next theorem makes some adjustment to the L-M quantizer so that it can be used to handle a set of correlated variables provided that the joint distribution is known.

Theorem 3.2.1

Let $\mathbf{X} = (x_1, x_2, \dots, x_p)'$ be a random vector with joint density $f(x_1, x_2, \dots, x_p)$ and $\mathbf{X}_{(-i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)'$, a vector from $\mathbf{X}$ with the i th component omitted. The quantizer minimizing the total mean square quantization error $E[\sum_{i=1}^p (x_i - \hat{x}_i)^2]$ is obtained by applying L-M quantizer on each component x_i with respect to the conditional density $f(x_i | \mathbf{X}_{(-i)})$.

Proof of Theorem 3.2.1

Define

$$\int d\mathbf{X}_{(-i)} = \overbrace{\int \cdots \int}^{p-1} dx_1, \dots, dx_{i-1}, dx_{i+1}, \dots, dx_p.$$

Then,

$$E\left[\sum_{i=1}^p (x_i - \hat{x}_i)^2\right]$$

$$\begin{aligned}
&= \sum_{i=1}^p \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_i - \hat{x}_i)^2 f(x_1, x_2, \dots, x_p) dx_1 dx_2 \dots dx_p \\
&= \sum_{i=1}^p \int \left[\int_{-\infty}^{\infty} (x_i - \hat{x}_i)^2 f(x_i | X_{(-i)}) dx_i \right] f(X_{(-i)}) dX_{(-i)}.
\end{aligned}$$

To minimize the above expression is equivalent to minimize

$$\int_{-\infty}^{\infty} (x_i - \hat{x}_i)^2 f(x_i | X_{(-i)}) dx_i,$$

for all i 's, which can be done by applying the L - M quantizer to each component x_i with respect to the conditional probability. This completes the proof.

This theorem is useful when the parameters in the admissible region of the model (3.2) are restricted by one another, i.e. what value a parameter can take depends on the values of the other parameters.

3.3 Limiting Distribution of the Prediction Residuals

Before quantizing the residuals, we need to examine their statistical character and figure out what factors have an effect on it. Among different sources, the noise is an important one causing prediction error. Moreover, since prediction is made by the quantized parameters, the prior and the quantizer chosen for the quantization of the parameters are also closely related to the accuracy of the prediction. If these factors are all determined, the residuals $\{e_i\}$ depend only on the total number of bits and the bit allocation.

Because of the intersample correlation, the exact distribution of $\{e_i\}$ is hard to obtain. In case the number of available bits for the quantization of the sample is

large and the L-M quantizer is used, it can be proved that $\{e_i\}$ converges to $\{\epsilon_i\}$ in distribution.

Lemma 3.3.1

Let X be a random variable with finite second moment and with $f(x)$ its probability density function with bounded range. Let $\hat{X}$ be the L-M quantizer for X . Then, $\hat{X} \rightarrow X$ in mean square when the quantization level v tends to infinity.

Proof of Lemma 3.3.1

By Lemma 3.2.2,

$$E[(X - \hat{X})^2] \leq E(X^2) < \infty.$$

Recall from the approximation (2.7),

$$E[(X - \hat{X})^2] = \frac{1}{12v^2} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^3 + o(v^{-2}) \rightarrow 0.$$

as $v \rightarrow \infty$.

In 3.1, we have mentioned that the three groups of elements in a block consist of the parameters, the first p observations and the last $n - p$ prediction residuals respectively. It will be proved in next section that the optimal bit allocation will be determined according to the variances of the components if they are uncorrelated. Since $\{e_i, i = 1, 2, \dots, p\}$ have common variance and so do $\{e_i, i = p + 1, p + 2, \dots, n\}$, we will assign the same numbers of bits to the components inside each group. We are now going to find the limit distribution of $\{e_i\}$ based on the above bit allocation.

Theorem 3.3.1

Let $\{\epsilon_i\}$ and $\{e_i\}$ be defined as in the previous section. Let $\{v_{\phi_j}\}$, v_1 and v_2 denote the quantization levels of $\{\phi_j\}$, $\{e_i, i = 1, 2, \dots, p\}$ and $\{e_i, i = p + 1, p + 2, \dots, n\}$ respectively. Suppose that L-M scheme is used for quantization of the parameters

and the residuals. Then, $\{e_i\}$ converges to $\{\epsilon_i\}$ in mean square if v_ϕ , v_1 and v_2 tend to infinity.

Proof of Theorem 3.3.1

For $i \leq p$, $e_i \equiv \epsilon_i$ by definition.

For $i > p$,

$$\begin{aligned}
 & E[(e_i - \epsilon_i)^2] \\
 &= E \left[\left(x_i - \hat{\Phi}' \hat{X}_{i-1} - x_i + \Phi' X_{i-1} \right)^2 \right] \\
 &= E \left[\left(\hat{\Phi}' \hat{X}_{i-1} - \hat{\Phi}' X_{i-1} + \hat{\Phi}' X_{i-1} - \Phi' X_{i-1} \right)^2 \right] \\
 &= E \left\{ \left[\hat{\Phi}' (X_{i-1} - \hat{X}_{i-1}) + (\hat{\Phi} - \Phi)' X_{i-1} \right]^2 \right\} \quad (3.4)
 \end{aligned}$$

Since $(a + b)^2 \leq 2a^2 + 2b^2$ for any real numbers a and b , the expression (3.4) is less than or equal to

$$\begin{aligned}
 & 2E \left[\hat{\Phi}' (X_{i-1} - \hat{X}_{i-1}) (X_{i-1} - \hat{X}_{i-1})' \hat{\Phi} \right] \\
 & + 2E \left[(\hat{\Phi} - \Phi)' X_{i-1} X_{i-1}' (\hat{\Phi} - \Phi) \right] \\
 &= 2E \left[\sum_{j=1}^p \sum_{k=1}^p (x_{i-j} - \hat{x}_{i-j})(x_{i-k} - \hat{x}_{i-k}) \phi_j \phi_k \right]
 \end{aligned}$$

$$+2E \left[\sum_{j=1}^p \sum_{k=1}^p x_{i-j} x_{i-k} (\phi_j - \hat{\phi}_j) (\phi_k - \hat{\phi}_k) \right] \quad (3.5)$$

Recall that $x_i - \hat{x}_i = e_i - \hat{e}_i$ and the admissible region of model (3.2) is bounded.

$$\begin{aligned} & E \left[\sum_{j=1}^p \sum_{k=1}^p |(x_{i-j} - \hat{x}_{i-j})(x_{i-k} - \hat{x}_{i-k}) \phi_j \phi_k| \right] \\ & \leq \sum_{j=1}^p \sum_{k=1}^p \sup(|\phi_j \phi_k|) E[(e_{i-j} - \hat{e}_{i-j})(e_{i-k} - \hat{e}_{i-k})] \\ & \leq \sum_{j=1}^p \sum_{k=1}^p \sup(|\phi_j \phi_k|) \sqrt{E[(e_{i-j} - \hat{e}_{i-j})^2]} \sqrt{E[(e_{i-k} - \hat{e}_{i-k})^2]} \\ & \rightarrow 0, \end{aligned}$$

as $\{v_{\phi_j}\}, v_1$ and v_2 tend to infinity by Lemma 3.3.1 and Cauchy-Schwarz inequality.

Since x has normal distribution, it has finite fourth moment. So

$$\begin{aligned} & E \left[\sum_{j=1}^p \sum_{k=1}^p |x_{i-j} x_{i-k} (\phi_j - \hat{\phi}_j)(\phi_k - \hat{\phi}_k)| \right] \\ & \leq \sum_{j=1}^p \sum_{k=1}^p 2 \sup(|\phi_k|) \sqrt{E[(x_{i-j} x_{i-k})^2]} \sqrt{E[(\phi_j - \hat{\phi}_j)^2]} \\ & \rightarrow 0, \end{aligned}$$

as $\{v_{\phi_j}\}, v_1$ and v_2 tend to infinity.

Theorem 3.3.1 showed that the prediction residual e_i tends to ϵ_i marginally in mean square. We now show that $\{\epsilon_i\}$ are asymptotically independent.

Theorem 3.3.2

Let $\{e_i\}$ be defined as in Theorem 3.3.1. Then, $\{e_i, i > p\}$ are asymptotically independent.

Proof of Theorem 3.3.2

Since $\{e_i\}$ have an approximately marginal normal distribution when the quantization level is large, it suffices to prove they are asymptotically uncorrelated.

$$\begin{aligned}
 & E |e_i e_j - \epsilon_i \epsilon_j| \\
 &= E |(e_i e_j - \epsilon_i \epsilon_j) + (\epsilon_i \epsilon_j - \epsilon_i \epsilon_j)| \\
 &\leq E |e_i(e_j - \epsilon_j)| + E |(\epsilon_i - \epsilon_i)\epsilon_j| \\
 &\leq \sqrt{E(e_i^2)}\sqrt{E[(e_j - \epsilon_j)^2]} + \sqrt{E[(\epsilon_i - \epsilon_i)^2]}\sqrt{E(\epsilon_j^2)} \\
 &\rightarrow 0,
 \end{aligned}$$

as $\{v_{\phi_j}\}, v_1$ and $v_2 \rightarrow \infty$. So

$$E(e_i e_j) \rightarrow E(\epsilon_i \epsilon_j) = 0$$

for $i, j = p+1, p+2, \dots, n$. This proves the asymptotic independence of $\{e_i, i > p\}$.

By Theorem 3.3.1 and 3.3.2, the residuals $\{e_i, i > p\}$ can be approximated by $\{\epsilon_i, i > p\}$ when the number of available bits is large. So we can handle $\{e_i, i > p\}$ as a set of i.i.d. random variables which provides us with great convenience in further

approach. But it is obvious that, at small quantization levels, the common variance of $\{e_i, i > p\}$, say σ_e^2 , is greater than that of $\{\epsilon_i, i > p\}$. The increment in the variance is caused by the quantization errors of the parameters as well as the quantization errors of the previous residuals. It is desirable to examine the influence of these two factors on the prediction error in order to find the optimal bit allocation.

By definition, the prediction error $\{e_i, i > p\}$ can be written as

$$\begin{aligned} e_i &= x_i - \hat{\Phi}'\hat{X}_{i-1} \\ &= x_i - \Phi'X_{i-1} + \Phi'X_{i-1} - \Phi'\hat{X}_{i-1} + \Phi'\hat{X}_{i-1} - \hat{\Phi}'\hat{X}_{i-1} \\ &= \epsilon_i + (\Phi - \hat{\Phi})'\hat{X}_{i-1} + \Phi'(X_{i-1} - \hat{X}_{i-1}). \end{aligned}$$

Then,

$$\begin{aligned} \sigma_e^2 &= E(e_i^2) \\ &= \sigma_\epsilon^2 + E[(\Phi - \hat{\Phi})'\hat{X}_{i-1}\hat{X}_{i-1}'(\Phi - \hat{\Phi})] \\ &\quad + E[\Phi'(X_{i-1} - \hat{X}_{i-1})(X_{i-1} - \hat{X}_{i-1})'\Phi] \\ &\quad + 2E[\epsilon_i(\Phi - \hat{\Phi})'\hat{X}_{i-1}] + 2E[\epsilon_i\Phi'(X_{i-1} - \hat{X}_{i-1})] \\ &\quad + 2E[(\Phi - \hat{\Phi})'\hat{X}_{i-1}(X_{i-1} - \hat{X}_{i-1})'\Phi]. \end{aligned}$$

The fourth and fifth term are zero since ϵ_i is independent of Φ and X_{i-1} . The last term is approximately zero by the results given by the next two theorems.

Theorem 3.3.3

Let $X_1, X_2, \dots, X_p$ be the first p unpredictable variables from model (3.2.1). Then,

$$E[(X_i - \hat{X}_i)\hat{X}_j] = 0, \quad i, j = 1, 2, \dots, p.$$

Proof of Theorem 3.3.3

If $i = j$, it follows the property (2) of the L - M quantizer directly.

Assume $i \neq j$. Without loss of generality, let $i > j$. Considering the conditional expectation on X_i given $X_k, k = i-1, i-2, \dots, 1$ under model (3.2), we have

$$\begin{aligned} E[X_i | X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2}, \dots, X_1 = x_1] \\ = -E[X_i | X_{i-1} = -x_{i-1}, X_{i-2} = -x_{i-2}, \dots, X_1 = -x_1]. \end{aligned} \quad (3.6)$$

By Lemma (3.2.1), we have $(-\hat{X}) = -\hat{X}$. So the conditional expectation of the quantization error is an odd function of $\{x_k, k = i-1, i-2, \dots, 1\}$, i.e.,

$$\begin{aligned} E[(X_i - \hat{X}_i) | X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2}, \dots, X_j = x_j] \\ = -E[(X_i - \hat{X}_i) | X_{i-1} = -x_{i-1}, X_{i-2} = -x_{i-2}, \dots, X_j = -x_j]. \end{aligned} \quad (3.7)$$

Since $(X_{i-1}, X_{i-2}, \dots, X_1)'$ is a multivariate normal random vector with mean zero,

$$E[(X_i - \hat{X}_i) | X_{i-1}, X_{i-2}, \dots, X_1] = 0.$$

So

$$\begin{aligned}
 & E[(X_i - \hat{X}_i)\hat{X}_j] \\
 &= E\{\hat{X}_j E[(X_i - \hat{X}_i)|X_{i-1}, X_{i-2}, \dots, X_1]\} \\
 &= 0
 \end{aligned}$$

for $i, j = 1, 2, \dots, p$.

Theorem 3.3.4

X_{i-1} and $\hat{X}_{i-1}$ are defined as above. Then,

$$E[\hat{X}_{i-1}(X_{i-1} - \hat{X}_{i-1})' | \Phi] \rightarrow 0, \quad (3.8)$$

as $\{v_{\phi_j}\}, v_1$ and v tend to infinity.

Proof of Theorem 3.3.4

The element in the j th row and k th column of the matrix $\hat{X}_{i-1}(X_{i-1} - \hat{X}_{i-1})'$ is $\hat{x}_{i-j}(x_{i-k} - \hat{x}_{i-k})$. We now prove that the mean of each element for fixed parameter Φ tends to zero.

If both $i-j$ and $i-k$ are less than p , the result follows by Theorem 3.3.3. Therefore, we need only consider the case that at least one of $i-j$ and $i-k$ is greater than or equal to p .

If $j = k$, note that $x_{i-j} - \hat{x}_{i-j} = e_{i-j} - \hat{e}_{i-j}$ and $\hat{x}_{i-j} = \hat{\Phi}'\hat{X}_{i-j-1} + \hat{e}_{i-j}$. Then,

$$\begin{aligned}
 & E[\hat{x}_{i-j}(x_{i-j} - \hat{x}_{i-j}) | \Phi] \\
 &= E[(\hat{\Phi}'\hat{X}_{i-j-1} + \hat{e}_{i-j})(e_{i-j} - \hat{e}_{i-j}) | \Phi] \\
 &= E[\hat{\Phi}'\hat{X}_{i-j-1}(e_{i-j} - \hat{e}_{i-j}) | \Phi] + E[\hat{e}_{i-j}(e_{i-j} - \hat{e}_{i-j}) | \Phi].
 \end{aligned}$$

By the properties of the L - M quantizer,

$$E[(e_{i-j} - \hat{e}_{i-j})|\Phi] = 0$$

$$E[\hat{e}_{i-j}(e_{i-j} - \hat{e}_{i-j})|\Phi] = 0,$$

Since e_{i-j} is asymptotically independent of $\hat{X}_{i-j-1}$ and $\hat{\Phi}$, the conditional mean vanishes in this case.

If $j > k$,

$$E[\hat{x}_{i-j}(x_{i-k} - \hat{x}_{i-k})|\Phi]$$

$$= E[\hat{x}_{i-j}(e_{i-k} - \hat{e}_{i-k})|\Phi]$$

$$\rightarrow 0$$

as $\{v_{\phi_j}\}, v_1$ and $v_2 \rightarrow \infty$ by the asymptotic independence of $\hat{x}_{i-j}$ and e_{i-k} .

If $j < k$,

$$E[\hat{x}_{i-j}(x_{i-k} - \hat{x}_{i-k})|\Phi]$$

$$= E[(\hat{\Phi}\hat{X}_{i-j-1} + \hat{e}_{i-j})(e_{i-k} - \hat{e}_{i-k})|\Phi]$$

$$= E\left\{\left[\sum_{l=1}^p f_l(\hat{\Phi})x_l + \sum_{l=p+1}^{i-k} g_l(\hat{\Phi})e_l\right](e_{i-k} - \hat{e}_{i-k})|\Phi\right\}$$

$$\rightarrow 0$$

as $\{v_{\phi_i}\}, v_1$ and $v_2 \rightarrow \infty$, where $f_1(\hat{\Phi})$ and $g_1(\hat{\Phi})$ are some functions obtained by recursively expressing x_i by the previous observations using model (3.2).

This completes the proof.

After eliminating the cross-product term, σ_e^2 may be written as

$$\begin{aligned}\sigma_e^2 &= \sigma_e^2 + E \left[(\Phi - \hat{\Phi})' \hat{X}_{i-1} \hat{X}_{i-1}' (\Phi - \hat{\Phi}) \right] \\ &\quad + E \left[\Phi' (X_{i-1} - \hat{X}_{i-1}) (X_{i-1} - \hat{X}_{i-1})' \Phi \right].\end{aligned}$$

The above expression still contains $\hat{X}_{i-1}$ whose distribution is hard to obtain. We have the following asymptotic result replacing $\hat{X}_{i-1}$ by X_{i-1} and $(X_{i-1} - \hat{X}_{i-1})$.

Lemma 3.3.2

Under the assumptions in model (3.2), we have

$$\begin{aligned}& E \left[(\Phi - \hat{\Phi})' \hat{X}_{i-1} \hat{X}_{i-1}' (\Phi - \hat{\Phi}) \right] \\ &= E \left[(\Phi - \hat{\Phi})' X_{i-1} X_{i-1}' (\Phi - \hat{\Phi}) \right] \\ &\quad - E \left[(\Phi - \hat{\Phi})' (X_{i-1} - \hat{X}_{i-1}) (X_{i-1} - \hat{X}_{i-1})' (\Phi - \hat{\Phi}) \right].\end{aligned}$$

Proof of Lemma 3.3.2

$$E \left[(\Phi - \hat{\Phi})' X_{i-1} X_{i-1}' (\Phi - \hat{\Phi}) \right]$$

$$\begin{aligned}
&= E \left[(\Phi - \hat{\Phi})' (X_{i-1} - \hat{X}_{i-1} + \hat{X}_{i-1}) (X_{i-1} - \hat{X}_{i-1} + \hat{X}_{i-1})' (\Phi - \hat{\Phi}) \right] \\
&= E \left[(\Phi - \hat{\Phi})' (X_{i-1} - \hat{X}_{i-1}) (X_{i-1} - \hat{X}_{i-1})' (\Phi - \hat{\Phi}) \right] \\
&\quad + E \left[(\Phi - \hat{\Phi})' \hat{X}_{i-1} \hat{X}_{i-1}' (\Phi - \hat{\Phi}) \right] \\
&\quad + 2E \left[(\Phi - \hat{\Phi})' (X_{i-1} - \hat{X}_{i-1}) \hat{X}_{i-1}' (\Phi - \hat{\Phi}) \right]
\end{aligned}$$

The last term vanishes by Theorem 3.3.4 and the result follows.

By Lemma 3.3.2, we have

$$\begin{aligned}
\sigma_e^2 &= \sigma_e^2 + E \left[(\Phi - \hat{\Phi})' X_{i-1} X_{i-1}' (\Phi - \hat{\Phi}) \right] \\
&\quad + E \left[\Phi' (X_{i-1} - \hat{X}_{i-1}) (X_{i-1} - \hat{X}_{i-1})' \Phi \right] \\
&\quad - E \left[(\Phi - \hat{\Phi})' (X_{i-1} - \hat{X}_{i-1}) (X_{i-1} - \hat{X}_{i-1})' (\Phi - \hat{\Phi}) \right]. \tag{3.9}
\end{aligned}$$

From (3.9), we can see the influence to the mean squared prediction error from different sources. The first term is the effect from the quantization errors of the parameters and the second term is that from the previous residuals. The last term is the interaction factor.

Using Lemma 3.3.2 again, the last two terms can be merged into one.

$$E \left[\Phi' (X_{i-1} - \hat{X}_{i-1}) (X_{i-1} - \hat{X}_{i-1})' \Phi \right]$$

$$\begin{aligned}
& -E \left[(\Phi - \hat{\Phi})'(X_{i-1} - \hat{X}_{i-1})(X_{i-1} - \hat{X}_{i-1})'(\Phi - \hat{\Phi}) \right] \\
& = E \left[\hat{\Phi}'(X_{i-1} - \hat{X}_{i-1})(X_{i-1} - \hat{X}_{i-1})'\hat{\Phi} \right].
\end{aligned}$$

So

$$\begin{aligned}
\sigma_e^2 &= \sigma_e^2 + E \left[(\Phi - \hat{\Phi})'X_{i-1}X_{i-1}'(\Phi - \hat{\Phi}) \right] \\
&+ E \left[\hat{\Phi}'(X_{i-1} - \hat{X}_{i-1})(X_{i-1} - \hat{X}_{i-1})'\hat{\Phi} \right]. \quad (3.10)
\end{aligned}$$

Equation (3.10) will be used to determine the optimal bit allocation.

3.4 Minimum MSE Bit Allocation

In the L-M scheme, the quantization error is a function of the quantization level if the input of the quantizer has fixed distribution. The approximation given by (2.6) reveals the association between the mean squared quantization error and the quantization levels.

Consider a set of uncorrelated random variables with unequal variances are to be quantized as a whole. Assuming that $\{x_i, i = 1, 2, \dots, n\}$ are independent normal random variables with variances $\{\sigma_i^2, i = 1, 2, \dots, n\}$ respectively, we are going to find the best choice, say b_i , of the number of bits assigned to x_i .

Huang and Schultheiss [15] studied this problem and obtained an optimal scheme for distributing bits among independent variables with unequal variances. Formula (2.12) given by Huang and Schultheiss can be used to approximate the optimal number of bits assigned to each variable.

Huang and Schultheiss' result can be obtained by a different approach which reveals some properties of the optimal bit allocation.

Lemma 3.4.1

Let x_1 and x_2 be real variables, and a_1, a_2, p and c be positive real constants. Then, subject to $x_1^p x_2 = c$,

$$f(x_1, x_2) = \frac{a_1}{x_1} + \frac{a_2}{x_2}$$

is minimized if and only if

$$\frac{x_2}{x_1} = \frac{pa_2}{a_1}.$$

Proof of Lemma 3.4.1

From the restriction, we have

$$x_2 = \frac{c}{x_1^p}$$

and

$$\frac{\partial x_2}{\partial x_1} = -\frac{px_2}{x_1}.$$

Hence,

$$\frac{\partial}{\partial x_1} \left(\frac{a_1}{x_1} + \frac{a_2}{x_2} \right) = -\frac{a_1}{x_1^2} + \frac{pa_2}{x_1 x_2}.$$

The first order partial derivative equals zero if and only if

$$\frac{x_2}{x_1} = \frac{pa_2}{a_1}.$$

It is easy to check that the second order partial derivative is positive, so the extreme value we found is indeed the minimum.

Theorem 3.4.1

Let $X_i \sim N(0, \sigma_i^2)$, $i = 1, 2, \dots, n$, be the input for the L-M quantizer with total of b bits available. Then when b is large, the asymptotic minimum MSE bit allocation will assign b_i bits to X_i such that

$$\frac{\sigma_1^2}{v_1^2} \simeq \frac{\sigma_2^2}{v_2^2} \simeq \dots \simeq \frac{\sigma_n^2}{v_n^2},$$

where $v_i = 2^{b_i}$ is the quantization level for X_i .

Proof of Theorem 3.4.1

Since $\sum_{i=1}^n b_i = b$, we have $\prod_{i=1}^n v_i = 2^{\sum_{i=1}^n b_i} = 2^b$. Recall the approximation formula for the quantization error (2.6) as well as the result given by Lemma 3.2.1.

$$E[(X_i - \hat{X}_i)^2] \simeq \frac{\sigma_i^2}{12v_i^2} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^3,$$

where $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is the standard normal density function.

The constant part contributed nothing in the procedure of minimization, so it can be removed from the expression of the total MSE. Now it suffices to minimize $\sum_{i=1}^n \frac{\sigma_i^2}{v_i^2}$ subject to $\prod_{i=1}^n v_i = \text{Constant}$.

In the case of $n = 2$: Write $a_i = \sigma_i^2$ and $x_i = v_i^2$ for $i = 1, 2$, then the problem is the immediate result of Lemma 3.4.1.

Assume the result holds for $n - 1$. By induction we now show it also holds for n .
Note that

$$\sum_{i=1}^{n-1} \frac{\sigma_i^2}{v_i^2} = (n-1) \frac{\sigma_1^2}{v_1^2}$$

and

$$\prod_{i=1}^n v_i = v_1 \left(\frac{\sigma_2^2}{\sigma_1^2} v_1 \right) \dots \left(\frac{\sigma_{n-1}^2}{\sigma_1^2} v_1 \right) v_n = \frac{\prod_{i=2}^{n-1} \sigma_i^2}{\sigma_1^{2(n-2)}} v_1^{n-1} v_n = c v_1^{n-1} v_n,$$

say, by our assumption. The problem is reduced to minimize

$$\frac{(n-1)\sigma_1^2}{v_1^2} + \frac{\sigma_n^2}{v_n^2}$$

subject to $v_1^{n-1} v_n = 2^s/c$.

Setting $a_1 = (n-1)\sigma_1^2$, $a_2 = \sigma_n^2$, $x_i = v_i^2$, for $i = 1, 2$, and $p = n-1$, the result follows by using Lemma 3.4.1 again.

Theorem 3.4.2

Suppose the total number of the available bits is fixed. Then, Huang and Schultheiss' formula (2.12) gives approximately the optimal bit assignment among independent normal random variables $X_1, X_2, \dots, X_n$ in the sense of attaining the minimum MSE with the L-M quantizer.

Proof of Theorem 3.4.2

By Huang and Schultheiss' formula (2.12), the quantization levels for X_i is

$$v_i = 2^{\frac{s}{n}} \frac{\sigma_{x_i}^2}{\left(\prod_{i=1}^n \sigma_{x_i}^2 \right)^{\frac{1}{2n}}}. \quad (3.11)$$

Hence,

$$\frac{v_i}{v_j} = \frac{\sigma_{x_i}^2}{\sigma_{x_j}^2}. \quad (3.12)$$

The result follows Theorem 3.4.1.

Suppose that we have totally s bits available for quantizing each block of the signal from an $AR(p)$ process, and m bits will be assigned to the parameters Φ . Let m_j be the number of bits used to quantize ϕ_j , so the quantization level for ϕ_j is $v_{\phi_j} = 2^{m_j}$. For the b bits left after assigning m bits to the parameters, let b_1 and b_2 be the numbers of bits assigned to the first p observations and the last $n - p$ prediction residuals respectively. Then, by Huang and Schultheiss' formula, b_1 and b_2 can be calculated by

$$\begin{aligned} b_1 &= \frac{b}{n} + \frac{1}{2} \log_2 \frac{E(\gamma_0) \sigma_\epsilon^2}{[E(\gamma_0)^p (\sigma_\epsilon^2)^p (\sigma_\epsilon^2)^{n-p}]^{\frac{1}{n}}} = \frac{b}{n} + \frac{1}{2} \log_2 [E(\gamma_0)]^{\frac{n-p}{n}} \\ b_2 &= \frac{b}{n} + \frac{1}{2} \log_2 \frac{\sigma_\epsilon^2}{[E(\gamma_0)^p (\sigma_\epsilon^2)^p (\sigma_\epsilon^2)^{n-p}]^{\frac{p}{n}}} = \frac{b}{n} + \frac{1}{2} \log_2 [E(\gamma_0)]^{-\frac{p}{n}} \end{aligned} \quad (3.13)$$

where $\gamma_0 = \sigma_x^2 / \sigma_\epsilon^2$ is a function of Φ , and the corresponding quantization levels are

$$\begin{aligned} v_1 &= 2^{b_1} = 2^{\frac{b}{n}} [E(\gamma_0)]^{\frac{n-p}{2n}} \\ v_2 &= 2^{b_2} = 2^{\frac{b}{n}} [E(\gamma_0)]^{-\frac{p}{2n}} \end{aligned} \quad (3.14)$$

respectively.

It should be pointed out that the above formula will overassign bits to $\{e_i, i = 1, 2, \dots, p\}$, since σ_e^2 is used instead of $\sigma_{e_i}^2$. This does not cause much difference when n is large. If n is small, it should be adjusted by using an iterative method. Use σ_e^2 as the initial value to calculate b_1 and b_2 from (3.13), then, use b_1 and b_2 to obtain the approximation of $\sigma_{e_i}^2$ derived in the next section. Repeat this procedure recursively until the new values of $\{b_i, i = 1, 2\}$ are close enough to the previous ones.

Note that $b = s - m$ and from (3.14), we have

$$\frac{v_1}{v_2} = [E(\gamma_0)]^{\frac{1}{2}}. \quad (3.15)$$

So after $\{v_{\phi_1}, v_{\phi_2}, \dots, v_{\phi_p}\}$ are chosen, v_1 and v_2 are determined by (3.15). If the prior of Φ and the variance of e_i are known, σ_e^2 is a function depending only on $\{v_{\phi_1}, v_{\phi_2}, \dots, v_{\phi_p}\}$.

3.5 Optimal Bit Allocation among Parameters

In the previous section, the minimum MSE bit allocation for quantizing a set of uncorrelated random variables was discussed and the procedure was used to determine the bit distribution among the prediction residuals. But the same procedure cannot be simply applied to determine the bit distribution among the parameters since we are facing a different situation. The purpose of the parameter quantization is to get a minimum MSE reconstruction for the original signal but not the parameters themselves. The parameters have different contribution to the reconstruction error and the bit allocation should be decided for this purpose.

Since the prior density of $\{\phi_1, \phi_2, \dots, \phi_p\}$ is defined on the admissible region of the AR(p) model, the domain of one parameter often depends on the values taken by

the others. So the parameters are generally not uncorrelated. In this case, Theorem 3.2.1 will be applied to get the optimal quantization.

We are now going to find the functional form of $E[(e_i - \hat{e}_i)^2]$, expressed as a function of quantization levels. To do so, we need the next two lemmas.

Lemma 3.5.1

Let x be a variable with density function $f(x)$. $\hat{x}$ is the L - M quantizer with quantization level v . Then, we have the next approximation based on large value of v .

$$\begin{aligned} & \int_{-\infty}^{\infty} (x - \hat{x})^2 g(x) f(x) dx \\ & \simeq \frac{1}{12v^2} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^2 \int_{-\infty}^{\infty} g(x) f^{\frac{1}{3}}(x) dx \end{aligned} \quad (3.16)$$

where $g(x)$ is a continuous real function.

Proof of Lemma 3.5.1

Let $-\infty = c_0 < c_1 < \dots < c_v = \infty$ be the end points of the quantization intervals and q_i be the quantized value in $(c_{i-1}, c_i]$. Define $\Delta_i = c_i - c_{i-1}$. Note that the integral $\int_{-\infty}^{\infty} g(x) dx$ can be approximated by $\sum_{i=1}^v g(q_i) \Delta_i$ if v is large. So

$$\begin{aligned} & \int_{-\infty}^{\infty} (x - \hat{x})^2 g(x) f(x) dx \\ & = \sum_{i=1}^v \int_{c_{i-1}}^{c_i} (x - q_i)^2 g(x) f(x) dx \end{aligned} \quad (3.17)$$

Note that $g(x) \simeq g(q_i)$, for $x \in \Delta_i$, which may be moved outside the integral. In the derivation of the mean square quantization error in Section 2.2, we know that

$$\int_{c_{i-1}}^{c_i} f^{\frac{1}{3}}(x) dx \equiv C, \text{ for } i = 1, 2, \dots, v,$$

where C is some constant. We have

$$\begin{aligned}
 & \sum_{i=1}^v g(q_i) \int_{c_{i-1}}^{c_i} (x - q_i)^2 f(x) dx \\
 & \simeq \sum_{i=1}^v g(q_i) \frac{1}{12v^3} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^3 \\
 & = \frac{1}{12v^3} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^2 \sum_{i=1}^v \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right] g(q_i) \\
 & \simeq \frac{1}{12v^3} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^2 \sum_{i=1}^v [v f^{\frac{1}{3}}(q_i) \Delta_i] g(q_i) \\
 & = \frac{1}{12v^2} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^2 \sum_{i=1}^v g(q_i) f^{\frac{1}{3}}(q_i) \Delta_i \\
 & \simeq \frac{1}{12v^2} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^2 \int_{-\infty}^{\infty} g(x) f^{\frac{1}{3}}(x) dx.
 \end{aligned}$$

In a special case when $g(x) \equiv 1$, the above expression becomes

$$\frac{1}{12v^2} \left[\int_{-\infty}^{\infty} f^{\frac{1}{3}}(x) dx \right]^3,$$

which gives the approximation for the mean square error.

Lemma 3.5.2

Let x be a variable with density function $f(x)$. $\hat{x}$ is the L - M quantizer with quantization level v . $g(x)$ is a continuous real function and $g'(x)$ is bounded. Then,

$$\int_{-\infty}^{\infty} (x - \hat{x}) g(x) f(x) dx = O(v^{-2}), \quad (3.18)$$

where $O(v^n)$ means at most the same order of magnitude as v^n .

Proof of Lemma 3.5.2

Let $\{c_i\}$ and $\{q_i\}$ be defined as in Lemma 3.5.1. By a Taylor expansion, $g(x)$ can be expressed, in each quantization interval, as

$$g(x) = g(q_i) + g'(q_i)(x - q_i) + o((x - q_i)).$$

By eliminating the higher order infinitesimal,

$$\begin{aligned} & \int_{-\infty}^{\infty} (x - \hat{x})g(x)f(x)dx \\ & \simeq \sum_{i=1}^v g(q_i) \int_{c_{i-1}}^{c_i} (x - q_i)f(x)dx + \sum_{i=1}^v g'(q_i) \int_{c_{i-1}}^{c_i} (x - q_i)^2 f(x)dx \\ & \leq \sup g'(x) \int_{-\infty}^{\infty} (x - \hat{x})^2 f(x)dx \end{aligned}$$

Note that the first term in the next to last expression is zero and the result follows from (2.7).

Theorem 3.5.1

If the quantization levels are large, we have the next approximation

$$E[(\Phi - \hat{\Phi})' X_{i-1} X_{i-1}' (\Phi - \hat{\Phi})] \simeq \sigma_e^2 \sum_{j=1}^p \frac{G_j}{v_{\phi_j}^2},$$

where

$$G_j = \frac{1}{12} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \pi_{j|j^c}^{\frac{1}{2}} d\phi_j \right]^2 \left[\int_{-\infty}^{\infty} \gamma_0 \pi_{j|j^c}^{\frac{1}{2}} d\phi_j \right] \pi_{j^c} d\Phi_{j^c}$$

$$\pi_{j|j^c} = \pi(\phi_j | \phi_1, \dots, \phi_{j-1}, \phi_{j+1}, \dots, \phi_p)$$

$$\pi_{j^c} = \pi(\phi_1, \dots, \phi_{j-1}, \phi_{j+1}, \dots, \phi_p).$$

Proof of Theorem 3.5.1

$$\begin{aligned}
 & E \left[(\Phi - \hat{\Phi})' X_{i-1} X_{i-1}' (\Phi - \hat{\Phi}) \right] \\
 &= E \left\{ E \left[(\Phi - \hat{\Phi})' X_{i-1} X_{i-1}' (\Phi - \hat{\Phi}) | \Phi \right] \right\} \\
 &= E \left[(\Phi - \hat{\Phi})' \Gamma \sigma_\epsilon^2 (\Phi - \hat{\Phi}) \right] \\
 &= \sigma_\epsilon^2 E \left[\sum_{j=1}^p \sum_{k=1}^p \gamma_{j-k} (\phi_j - \hat{\phi}_j) (\phi_k - \hat{\phi}_k) \right], \tag{3.19}
 \end{aligned}$$

where

$$\Gamma = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & & & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{pmatrix}.$$

is such that $E[(X_{i-1} X_{i-1}') | \Phi] = \Gamma \sigma_\epsilon^2$.

Let $-\infty = c_{j0} < c_{j1} < \dots < c_{jv_{\phi_j}} = \infty$ be the end points of the quantization intervals of ϕ_j , $\Delta_{jl} = c_{jl} - c_{j,l-1}$, $l = 1, 2, \dots, v_{\phi_j}$ and q_{jl} be the quantized value in Δ_{jl} . We now prove that the terms in (3.19) with index $j \neq k$ are higher order infinitesimal than that with $j = k$.

If $j \neq k$,

$$E \left[\gamma_{j-k} (\phi_j - \hat{\phi}_j) (\phi_k - \hat{\phi}_k) \right]$$

$$\begin{aligned}
&= E \left\{ E \left[\gamma_{j-k}(\phi_j - \hat{\phi}_j)(\phi_k - \hat{\phi}_k) | \Phi_{(-j)} \right] \right\} \\
&= E \left\{ (\phi_k - \hat{\phi}_k) E \left[\gamma_{j-k}(\phi_j - \hat{\phi}_j) | \Phi_{(-j)} \right] \right\} \quad (3.20)
\end{aligned}$$

Note that $E \left[\gamma_{j-k}(\phi_j - \hat{\phi}_j) | \Phi_{(-j)} \right] = O(v_{\phi_j}^{-2})$. Using Lemma (3.5.2) again, we have

$$E \left[\gamma_{j-k}(\phi_j - \hat{\phi}_j)(\phi_k - \hat{\phi}_k) \right] = O(v_{\phi_j}^{-2} v_{\phi_k}^{-2}) = o(v_{\phi_l}^{-2}).$$

for $l = j, k$. So we can consider only the terms with $j = k$.

$$\begin{aligned}
&E \left[\gamma_0(\phi_j - \hat{\phi}_j)^2 \right] \\
&= E \left\{ E \left[\gamma_0(\phi_j - \hat{\phi}_j)^2 | \phi_1, \dots, \phi_{j-1}, \phi_{j+1}, \dots, \phi_p \right] \right\} \\
&\simeq \int_{-\infty}^{\infty} \frac{1}{12v_{\phi_j}^2} \left[\int_{-\infty}^{\infty} \pi_{j|j^c}^{\frac{1}{2}} d\phi_j \right]^2 \left[\int_{-\infty}^{\infty} \gamma_0 \pi_{j|j^c}^{\frac{1}{2}} d\phi_j \right] d\Phi_{(-j)} \\
&= \frac{G_j}{v_{\phi_j}^2}.
\end{aligned}$$

So by eliminating the higher order infinitesimal ,

$$E \left[(\Phi - \hat{\Phi})' X_{i-1} X_{i-1}' (\Phi - \hat{\Phi}) \right] \simeq \sigma_\epsilon^2 \sum_{j=1}^p \frac{G_i}{v_{\phi_j}^2}.$$

Theorem 3.5.2

$$E \left[\hat{\Phi}'(X_{i-1} - \hat{X}_{i-1})(X_{i-1} - \hat{X}_{i-1})' \hat{\Phi} \right]$$

$$\simeq \sum_{j=1}^p \left[E(\phi_j^2) - \frac{D_j}{v_{\phi_j}^2} \right] \frac{I \sigma_e^2}{v_2^2},$$

where

$$D_j = \frac{1}{12} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \pi_{j|j^c}^{\frac{1}{2}} d\phi_j \right]^3 d\Phi_{(-j)},$$

$$I = \frac{1}{12} \left[\int_{-\infty}^{\infty} f^{\frac{1}{2}}(x) dx \right]^3$$

and $f(x)$ is the standard normal density function.

Proof of Theorem 3.5.2

$$E \left[\hat{\Phi}'(X_{i-1} - \hat{X}_{i-1})(X_{i-1} - \hat{X}_{i-1})' \hat{\Phi} \right]$$

$$= E \left[\sum_{j=1}^p \sum_{k=1}^p (x_{i-j} - \hat{x}_{i-j})(x_{i-k} - \hat{x}_{i-k}) \hat{\phi}_j \hat{\phi}_k \right]$$

$$= E \left[\sum_{j=1}^p \sum_{k=1}^p (e_{i-j} - \hat{e}_{i-j})(e_{i-k} - \hat{e}_{i-k}) \hat{\phi}_j \hat{\phi}_k \right].$$

Since all e_i 's have same quantization error by Theorem 3.4.2, we can ignore whether $i \leq p$ or $i > p$.

If $j \neq k$, then,

$$E \left[(e_{i-j} - \hat{e}_{i-j})(e_{i-k} - \hat{e}_{i-k})\hat{\phi}_j\hat{\phi}_k \right] \simeq 0$$

since e_i 's are asymptotically independent of ϕ_i 's and independent of one another.

If $j = k$, then,

$$E \left[(e_{i-j} - \hat{e}_{i-j})^2 \hat{\phi}_j^2 \right]$$

$$\simeq E \left[(e_{i-j} - \hat{e}_{i-j})^2 \right] E \left[\hat{\phi}_j^2 \right],$$

because of the independence of e_i 's and ϕ_i 's. By applying Lemma (3.2.2), we get

$$\begin{aligned} E \left[\hat{\phi}_j^2 \right] &= E \left[\phi_j^2 \right] - E \left[(\phi_j - \hat{\phi}_j)^2 \right] \\ &= E \left[\phi_j^2 \right] - \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} (\phi_j - \hat{\phi}_j)^2 \pi_{j|j^c} d\phi_i \right] \pi_{j^c} d\Phi_{(-j)} \\ &\simeq E \left[\phi_j^2 \right] - \int_{-\infty}^{\infty} \frac{1}{12v_{\phi_j}^2} \left[\int_{-\infty}^{\infty} \pi_{j|j^c}^{\frac{1}{3}} d\phi_i \right]^3 \pi_{j^c} d\Phi_{(-j)} \\ &= E \left[\phi_j^2 \right] - \frac{D_j}{v_{\phi_j}^2}. \end{aligned}$$

and

$$E[(e_{i-j} - \hat{e}_{i-j})^2] \simeq \frac{I\sigma_c^2}{v_2^2}$$

by Lemma 3.5.1. So

$$E \left[\hat{\Phi}'(X_{i-1} - \hat{X}_{i-1})(X_{i-1} - \hat{X}_{i-1})' \hat{\Phi} \right]$$

$$\simeq \sum_{j=1}^p \left[E(\phi_j^2) - \frac{D_j}{v_{\phi_j}^2} \right] \frac{I\sigma_\epsilon^2}{v_2^2}.$$

Combining Theorem 3.5.1 and 3.5.2, the expression of σ_ϵ^2 given in (3.10) can be written as

$$\sigma_\epsilon^2 \simeq \sigma_\epsilon^2 + \sigma_\epsilon^2 \sum_{j=1}^p \frac{G_j}{v_{\phi_j}^2} + \sum_{j=1}^p \left[E(\phi_j^2) - \frac{D_j}{v_{\phi_j}^2} \right] \frac{I\sigma_\epsilon^2}{v_2^2}. \quad (3.21)$$

The above equation implies that σ_ϵ^2 is a multiple of σ_ϵ^2 for fixed quantization levels.

We can write

$$\sigma_\epsilon^2 = C(v_{\phi_1}, v_{\phi_2}, \dots, v_{\phi_p}, v_1, v_2) \sigma_\epsilon^2, \quad (3.22)$$

where

$$C(v_{\phi_1}, v_{\phi_2}, \dots, v_{\phi_p}, v_1, v_2) = \frac{1 + \sum_{j=1}^p G_j v_{\phi_j}^{-2}}{1 - \sum_{j=1}^p \left[E(\phi_j^2) - D_j v_{\phi_j}^{-2} \right] I v_2^{-2}} \quad (3.23)$$

By Lemma 3.2.1, the mean squared quantization error has the expression

$$E[(e_i - \hat{e}_i)^2] = \frac{1 + \sum_{j=1}^p G_j v_{\phi_j}^{-2}}{v_2^2 I^{-1} - \sum_{j=1}^p \left[E(\phi_j^2) - D_j v_{\phi_j}^{-2} \right]} \quad (3.24)$$

We can now write the total mean squared error that we want to minimize as

$$\begin{aligned}
 & E \sum_{i=1}^n [(e_i - \hat{e}_i)^2] \\
 &= \sum_{i=1}^p E[(e_i - \hat{e}_i)^2] + \sum_{i=p}^n E[(e_i - \hat{e}_i)^2] \\
 &\simeq n\sigma_\epsilon^2 \frac{1 + \sum_{j=1}^p G_j v_{\phi_j}^{-2}}{v_2^2 I^{-1} - \sum_{j=1}^p [E(\phi_j^2) - D_j v_{\phi_j}^{-2}]} \\
 &= n\sigma_\epsilon^2 H \tag{3.25}
 \end{aligned}$$

say. The approximation above is based on a direct result from Theorem 3.4.1 that the quantization errors for all the components in a vector will be same if Huang-Schultheiss bit allocation scheme is used.

Since we are interested in the average values of the quantization levels, they will be treated as continuous variables although it is not the case in practice. So the best choice of $\{v_{\phi_j}, j = 1, 2, \dots, p\}$ can be determined by differentiating the expression (3.25) with respect to $v_{\phi_j}^2$ and setting the derivative to zero. Note that

$$v_{\phi_j} = 2^{m_j}$$

$$v_2 = 2^{\frac{s - \sum_{j=1}^p m_j}{\frac{j=1}{n}} [E(\gamma_0)]^{-\frac{p}{2n}}}$$

$$= v_{\phi_j}^{-\frac{1}{n}} 2^{\frac{s - \sum_{k \neq j}^p m_j}{n}} [E(\gamma_0)]^{-\frac{p}{2n}}.$$

so

$$\frac{\partial v_2^2}{\partial v_{\phi_j}^2} = -\frac{v_2^2}{n v_{\phi_j}^2}.$$

We have

$$\frac{\partial H}{\partial v_{\phi_j}^2} = \frac{U_j + V_j}{W^2},$$

where

$$W = v_2^2 I^{-1} - \sum_{j=1}^p [E(\phi_j^2) - D_j v_{\phi_j}^{-2}]$$

$$U_j = -\frac{G_j}{v_{\phi_j}^4} W$$

$$V_j = \left(1 + \sum_{j=1}^p \frac{G_j}{v_{\phi_j}^2}\right) \left(\frac{v_2^2}{n I v_{\phi_j}^2} + \frac{D_j}{v_{\phi_j}^4}\right).$$

The optimal v_{ϕ_j} 's are obtained by solving the equations

$$-\frac{G_j}{v_{\phi_j}^4} \frac{(1 + \sum_{j=1}^p G_j v_{\phi_j}^{-2})}{\{v_2^2 I^{-1} - \sum_{j=1}^p [E(\phi_j^2) - D_j v_{\phi_j}^{-2}]\}} + \left(1 + \sum_{j=1}^p \frac{G_j}{v_{\phi_j}^2}\right) \left(\frac{v_2^2}{n I v_{\phi_j}^2} + \frac{D_j}{v_{\phi_j}^4}\right)$$

$$= 0,$$

for $j = 1, 2, \dots, p$ simultaneously subject to $\frac{v_1}{v_2} = [E(\gamma_0)]^{\frac{1}{2}}$ and $m + b = s$.

In 3.4, we have mentioned that when the number of available bits is small, the value σ_e^2 can be used to adjust the bit allocation between v_1 and v_2 . Conversely, we can calculate σ_e^2 again from the adjusted value of v_2 . It can be proved that we can get the optimal choice for v_2 by repeating this procedure. For fixed $\{v_{\phi_j}\}$, let $v_2^{(1)}$ be the initial value for v_2 obtained from (3.10) using σ_e^2 . Note that for fixed $\{v_{\phi_j}\}$, $v_2^{(1)}$ is the smallest quantization level we could possibly assign to the residuals since $\sigma_e^2 \geq \sigma_e^2$. Let $\sigma_e^{2(1)}$ be the value calculated from (3.22) using the initial $v_2^{(1)}$. Note that the expression of σ_e^2 in (3.22) is a decreasing function of v_2 . So $\sigma_e^{2(1)}$ is the largest possible value for σ_e^2 if Huang-Schultheiss bit allocation scheme is used. Using (3.10) again, $v_2^{(2)}$ obtained from $\sigma_e^{2(1)}$ will attain the maximum for v_2 and will lead to the smallest value $\sigma_e^{2(2)}$ using the calculated σ_e^2 . But note that $\sigma_e^{2(2)} > \sigma_e^2$ unless the quantization levels $\{v_{\phi_j}\}$ and v_2 tend to infinity. Repeating this procedure recursively, we have

$$v_2^{(1)} < v_2^{(3)} < \dots < v_2^{(2n+1)} < \dots$$

$$v_2^{(2)} > v_2^{(4)} > \dots > v_2 > (2) > \dots$$

So $v_2^{(n)}$ will converge and its limit will be the optimal choice for v_2 .

3.6 Simulation and Discussion

A computer simulation was run to evaluate the performance of the Bayesian adaptive DPCM scheme discussed above. The data used in the simulation were from

an AR(1) model. Since there is only one parameter in this model, we will omit the index in the following discussion.

First, we evaluated the goodness of the asymptotic approximation at low quantization levels. In the simulation, the parameter ϕ was generated from a uniform distribution (Table 3.1) and a distribution with linear density function (Table 3.2). Both distributions were defined on the interval $[0, 1 - \delta]$ with $\delta = 10^{-4}$. Processes of length $n = 20, 40, 60, 80, 100$ with normally distributed white noise with zero mean and unit variance were generated using the parameter ϕ . The parameter and the prediction residuals were then quantized using the Bayesian adaptive DPCM scheme with s/n , the average bits for each observation, equals to 3, 4, and 5, respectively. The theoretical mean square quantization error $E[(e_i - \hat{e}_i)^2]$ was calculated by using the approximation formula (3.25) and the estimate $\hat{E}[(e_i - \hat{e}_i)^2]$ was obtained by the average value of 1000 simulated series. Since in practice the bit number and the quantization level take only integer values, the optimal values are rounded up to the nearest integer. The result is listed in Table 3.1 (for uniform prior) and Table 3.2 (for linear density prior).

From Table 3.1 and Table 3.2, we can see that the MSE from the simulation results are very close to that obtained by the asymptotic approximation for $m \geq 2$, and the difference turns smaller when the total number of available bits increases. The optimal bit assignments from the simulation mostly coincide with that from the theoretical approximation with the farthest departure of one bit. Since the mean squared quantization error changes very little in the neighborhood of the optimal v_ϕ , this slight departure will not make big difference in the general performance of this scheme.

We are also interested in examining how the mean squared quantization error is affected by the variation of the other factors. For discussion convenience, we rewrite

the expressions (3.23) and (3.24) subject to AR(1) model as follows

$$G(v_\phi, v_1, v_2) = \frac{v_\phi^2 + G}{v_\phi^2 - [E(\phi^2)v_\phi^2 - D]Iv_2^{-2}}, \quad (3.26)$$

$$E[(e_i - \hat{e}_i)^2] \simeq \frac{(v_\phi^2 + G)\sigma_\epsilon^2}{v_2^2 v_\phi^2 I^{-1} - E(\phi^2)v_\phi^2 + D}. \quad (3.27)$$

Among several factors which might affect the optimal choice of v_ϕ , the sample size n has the most significant influence. Table 3.3 gives us the optimal choices of m against the series length n with average bit rates 3, 4, and 8. The optimal values of v_ϕ are also listed since it has smaller rounding error. We can see that the optimal v_ϕ is an increasing function of n , which means we should assign more bits for the quantization of the parameter if the sample size becomes larger. This conclusion seems quite reasonable.

It seems surprising that the average bit rate, when it is not so small (≥ 3), has little effect on the optimal choice of v_ϕ . An explanation can be obtained from (3.27). The constants D and G in the expression are generally small compared with the values of the quantization levels v_ϕ, v_1 and v_2 . Note that D and G depend only on the prior density assigned to the parameters and in our simulation, $D = 0.0833$ and $G = 0.4126$ for the uniform prior and $D = 0.0703$ and $G = 0.4323$ for the linear prior. So when the quantization levels are large, these constants can be eliminated from the expression (3.27). So we have

$$E[(e_i - \hat{e}_i)^2] \simeq \frac{I\sigma_\epsilon^2}{v_2^2 - IE(\phi^2)}. \quad (3.28)$$

From (3.28), it is clear that the quantization error depends on the parameter quantization only through v_2 which is fairly stable under the change of v_ϕ .

The variance of the white noise σ_e^2 has no effect on the optimal bit allocation. Equation (3.27) shows that the mean square quantization error is indeed proportional to σ_e^2 . But in (3.26), $C(v_\phi, v_1, v_2)$, from which the optimal value of v_ϕ is determined, is independent of σ_e^2 . This gives us an explanation why the optimal design holds for different values of σ_e^2 .

The value $E(\phi^2)$ is usually small because of the restriction of admissibility. For instance, the admissible regions for AR(1) and AR(2) models are $\{\phi : |\phi| < 1\}$ and $\{(\phi_1, \phi_2) : \phi_2 + \phi_1 < 1, \phi_2 - \phi_1 < 1, \text{ and } |\phi_2| < 1\}$, respectively. We have $E(\phi^2) \leq 1$ for the AR(1) model and $E(\phi_1^2) \leq 4$ and $E(\phi_2^2) \leq 1$ for the AR(2) model. The upper bounds are very unlikely to be attained, and for most practically used prior distributions, the variance is far lower than the upper bound. In contrast, the square of the quantization level is usually a fairly large number. Hence, the quantization error can be approximated by

$$E[(e_i - \hat{e}_i)^2] \simeq \frac{I\sigma_e^2}{v_2^2}, \quad (3.29)$$

which is the limit of the quantization error for the white noise.

Comparison was made with the conventional DPCM and Cummiskey's adaptive scheme. Two types of nonstationary time series were used in the simulation. The first type of AR(1) series was composed by several subsequences with different lengths which was decided by a random variable from a Poisson distribution with mean λ . The parameter ϕ was from a uniform distribution. $\phi \sim U[0, 1 - \delta]$ with $\delta = 10^{-4}$. The second type of nonstationary series was generated by using a time-variate parameter. The initial ϕ had a uniform distribution on the interval $[0, 1 - \delta]$ and a small noise $\Delta\phi \sim U(-0.1, 0.1)$ was added to it after each observation. $\phi + \Delta\phi$ is truncated by the upper and lower bounds 0 and $1 - \delta$ if it is out of range. In the simulation, both types of series with length $n = 1200$ was divided into certain number of blocks with equal

size l and each block was treated individually. In DPCM, the prediction was made using a constant $\hat{\phi} = E(\phi)$ throughout the whole block without any adjustment, so no bit was needed to quantize the parameter. In the Bayesian approach, ϕ was estimated, and then, prediction and quantization were performed within the block. In Cummiskey's scheme, the first 10 observations were used to get the initial estimation for the parameter ϕ . Each estimate was used only once for the prediction for the next value and after that, the estimate was updated based on the new observation according to

$$\Delta\phi = \frac{K e_i}{x_{i-1}}. \quad (3.30)$$

The best adaptive rate K was found to be 0.1 by comparing different rates. The first 10 data were quantized directly while others were predicted and the residuals were quantized. The mean squared quantization errors using different schemes are listed in the columns in Table 3.4 (first type) and Table 3.5 (second type) with different choices of bit rates and block lengths in each row.

Results show that the Bayesian scheme works much better in the simulated cases. From (3.30), we can see that the adaptation in the Cummiskey's scheme depends only on one previous observation as well as the current prediction error, so the process can be very unstable. An observation near zero will cause a big jerk even the prediction is not so bad. Specially, if the jerk leads to the wrong direction, it is very difficult to pull it back. Another weakness of the Cummiskey's adaptation is its inability to handle the outlier. The increment in the adaptation is proportional to the prediction error, so any bad prediction will make the further adaptation even worse. Furthermore, the parameter adaptation based on the current prediction error may not work well to predict the next observation. In the Bayesian adaptive scheme, the prediction is made by using the data of the whole block, therefore, it is not so sensitive to outliers. The cost of a few bits to quantize the parameter gets big payoff in reducing the

prediction errors. This makes the Bayesian scheme superior in dealing with highly nonstationary signals.

Table 3.1. Mean squared quantization errors from asymptotic approximation ($E[(e_i - \hat{e}_i)^2]$) and simulation ($\hat{E}[(e_i - \hat{e}_i)^2]$). $\phi \sim U(0, 1 - \delta)$, $\delta = 10^{-4}$.

$n = 20$

m	$s/n=3$		$s/n=4$		$s/n=5$	
	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$
0	0.05152	0.06695	0.01438	0.01814	0.00382	0.00545
1	0.04294	0.04404	0.01199	0.01348	0.00319	0.00339
2	0.04252	0.04249	0.01190	0.01221	0.00317	0.00317
3	0.04440	0.04380	0.01246	0.01268	0.00333	0.00320
4	0.04700	0.04575	0.01324	0.01339	0.00354	0.00349
5	0.04993	0.04976	0.01411	0.01397	0.00378	0.00371
6	0.05308	0.05253	0.01504	0.01505	0.00404	0.00402
7	0.05644	0.05500	0.01604	0.01587	0.00432	0.00433
8	0.05999	0.06008	0.01711	0.01720	0.00461	0.00449
9	0.06377	0.06312	0.01825	0.01810	0.00493	0.00478
10	0.06776	0.06833	0.01945	0.01918	0.00527	0.00509

s = total number of bits available, m = number of bits assigned to ϕ , n = length of block.

Table 3.1. Continued. $n = 40$

m	$s/n=3$			$s/n=4$		$s/n=5$	
	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$
0	0.04970	0.07227	0.01384	0.01702	0.00367	0.00463	
1	0.04016	0.04433	0.01118	0.01158	0.00296	0.00307	
2	0.03855	0.04004	0.01074	0.01045	0.00285	0.00277	
3	0.03903	0.04019	0.01089	0.01060	0.00289	0.00277	
4	0.04007	0.04094	0.01120	0.01091	0.00298	0.00282	
5	0.04128	0.04232	0.01155	0.01130	0.00308	0.00298	
6	0.04257	0.04356	0.01193	0.01150	0.00318	0.00305	
7	0.04391	0.04465	0.01232	0.01187	0.00329	0.00317	
8	0.04529	0.04634	0.01273	0.01229	0.00340	0.00330	
9	0.04671	0.04825	0.01315	0.01284	0.00351	0.00335	
10	0.04817	0.04997	0.01358	0.01322	0.00363	0.00346	

Table 3.1. Continued. $n = 60$

m	$s/n=3$		$s/n=4$		$s/n=5$	
	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$
0	0.04911	0.06969	0.01367	0.01844	0.00362	0.00445
1	0.03927	0.04380	0.01092	0.01164	0.00289	0.00299
2	0.03731	0.03910	0.01038	0.01027	0.00275	0.00271
3	0.03738	0.03867	0.01041	0.01021	0.00276	0.00266
4	0.03798	0.03962	0.01059	0.01039	0.00281	0.00274
5	0.03873	0.04013	0.01080	0.01050	0.00287	0.00276
6	0.03953	0.04074	0.01104	0.01068	0.00293	0.00280
7	0.04036	0.04190	0.01128	0.01090	0.00300	0.00285
8	0.04120	0.04302	0.01153	0.01116	0.00307	0.00289
9	0.04206	0.04362	0.01178	0.01133	0.00314	0.00298
10	0.04294	0.04433	0.01204	0.01155	0.00321	0.00308

Table 3.1. Continued. $n = 80$

m	$s/n=3$		$s/n=4$		$s/n=5$	
	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$
0	0.04882	0.06756	0.01358	0.01937	0.00360	0.00463
1	0.03883	0.04230	0.01079	0.01163	0.00286	0.00303
2	0.03670	0.03760	0.01020	0.01004	0.00270	0.00269
3	0.03658	0.03706	0.01017	0.00975	0.00269	0.00269
4	0.03698	0.03754	0.01029	0.00978	0.00273	0.00270
5	0.03751	0.03810	0.01045	0.01004	0.00277	0.00274
6	0.03809	0.03866	0.01062	0.01015	0.00282	0.00278
7	0.03869	0.03924	0.01079	0.01029	0.00287	0.00282
8	0.03929	0.03973	0.01097	0.01042	0.00291	0.00289
9	0.03991	0.04014	0.01115	0.01063	0.00296	0.00296
10	0.04053	0.04088	0.01133	0.01075	0.00301	0.00300

Table 3.1. Continued. $n = 100$

m	$s/n=3$		$s/n=4$		$s/n=5$	
	$E[(e_r - \hat{e}_r)^2]$	$\hat{E}[(e_r - \hat{e}_r)^2]$	$E[(e_r - \hat{e}_r)^2]$	$\hat{E}[(e_r - \hat{e}_r)^2]$	$E[(e_r - \hat{e}_r)^2]$	$\hat{E}[(e_r - \hat{e}_r)^2]$
0	0.04864	0.07389	0.01353	0.01782	0.00359	0.00464
1	0.03857	0.04423	0.01072	0.01138	0.00284	0.00301
2	0.03634	0.03789	0.01010	0.01018	0.00268	0.00266
3	0.03611	0.03699	0.01004	0.00995	0.00266	0.00260
4	0.03639	0.03750	0.01012	0.00996	0.00268	0.00261
5	0.03680	0.03753	0.01024	0.01005	0.00272	0.00263
6	0.03725	0.03829	0.01037	0.01010	0.00275	0.00267
7	0.03772	0.03861	0.01051	0.01033	0.00279	0.00272
8	0.03819	0.03889	0.01065	0.01043	0.00283	0.00273
9	0.03867	0.03935	0.01079	0.01067	0.00287	0.00278
10	0.03915	0.03986	0.01093	0.01083	0.00290	0.00282

Table 3.2. Mean squared quantization errors from asymptotic approximation and simulation. $\pi(\phi) = (1 - \delta)^{-2} 2\phi$, $\delta = 10^{-4}$.
 $n = 20$

m	$s/n=3$		$s/n=4$		$s/n=5$	
	$E[(e_r - \hat{e}_r)^2]$	$\hat{E}[(e_r - \hat{e}_r)^2]$	$E[(e_r - \hat{e}_r)^2]$	$\hat{E}[(e_r - \hat{e}_r)^2]$	$E[(e_r - \hat{e}_r)^2]$	$\hat{E}[(e_r - \hat{e}_r)^2]$
0	0.05398	0.05085	0.01498	0.01334	0.00397	0.00365
1	0.04458	0.04379	0.01238	0.01107	0.00329	0.00308
2	0.04403	0.04225	0.01225	0.01140	0.00326	0.00304
3	0.04595	0.04427	0.01282	0.01175	0.00342	0.00315
4	0.04867	0.04580	0.01361	0.01269	0.00363	0.00339
5	0.05174	0.04919	0.01451	0.01318	0.00388	0.00355
6	0.05505	0.05253	0.01547	0.01402	0.00415	0.00380
7	0.05857	0.05693	0.01651	0.01472	0.00443	0.00406
8	0.06232	0.05987	0.01761	0.01602	0.00474	0.00432

For notation, see Table 3.1.

Table 3.2. Continued. $n = 40$

m	$s/n=3$		$s/n=4$		$s/n=5$	
	$E[(e_r - \hat{e}_r)^2]$	$\hat{E}[(e_r - \hat{e}_r)^2]$	$E[(e_r - \hat{e}_r)^2]$	$\hat{E}[(e_r - \hat{e}_r)^2]$	$E[(e_r - \hat{e}_r)^2]$	$\hat{E}[(e_r - \hat{e}_r)^2]$
0	0.05143	0.05095	0.01425	0.01254	0.00377	0.00333
1	0.04116	0.04070	0.01139	0.01063	0.00302	0.00280
2	0.03939	0.03987	0.01091	0.01023	0.00289	0.00268
3	0.03985	0.04065	0.01105	0.01023	0.00293	0.00270
4	0.04092	0.04183	0.01136	0.01063	0.00302	0.00284
5	0.04217	0.04276	0.01173	0.01108	0.00312	0.00290
6	0.04350	0.04381	0.01211	0.01127	0.00322	0.00303
7	0.04488	0.04542	0.01251	0.01168	0.00333	0.00309
8	0.04630	0.04702	0.01292	0.01207	0.00345	0.00322

Table 3.2. Continued. $n = 60$

m	$s/n=3$		$s/n=4$		$s/n=5$	
	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$
0	0.05060	0.05019	0.01401	0.01372	0.00371	0.00336
1	0.04007	0.04658	0.01108	0.01126	0.00293	0.00270
2	0.03795	0.03797	0.01050	0.01055	0.00278	0.00265
3	0.03800	0.03819	0.01052	0.01069	0.00279	0.00264
4	0.03861	0.03912	0.01070	0.01086	0.00284	0.00270
5	0.03938	0.03984	0.01092	0.01113	0.00290	0.00277
6	0.04020	0.04079	0.01116	0.01142	0.00296	0.00282
7	0.04105	0.04121	0.01140	0.01154	0.00303	0.00283
8	0.04191	0.04231	0.01165	0.01185	0.00310	0.00290

Table 3.2. Continued. $n = 80$

m	$s/n=3$		$s/n=4$		$s/n=5$	
	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$
0	0.05020	0.04863	0.01389	0.01332	0.00368	0.00345
1	0.03954	0.03937	0.01093	0.01043	0.00289	0.00288
2	0.03725	0.03709	0.01030	0.00981	0.00273	0.00271
3	0.03710	0.03735	0.01026	0.00970	0.00272	0.00268
4	0.03750	0.03779	0.01038	0.00979	0.00275	0.00272
5	0.03805	0.03838	0.01054	0.00992	0.00279	0.00279
6	0.03864	0.03904	0.01071	0.01012	0.00284	0.00282
7	0.03925	0.03942	0.01088	0.01025	0.00289	0.00288
8	0.03987	0.04003	0.01106	0.01045	0.00294	0.00290

Table 3.2. Continued. $n = 100$

m	$s/n=3$		$s/n=4$		$s/n=5$	
	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$	$E[(e_i - \hat{e}_i)^2]$	$\hat{E}[(e_i - \hat{e}_i)^2]$
0	0.04995	0.05113	0.01382	0.01301	0.00366	0.00343
1	0.03923	0.03993	0.01084	0.01052	0.00287	0.00273
2	0.03684	0.03708	0.01018	0.00999	0.00269	0.00256
3	0.03658	0.03674	0.01011	0.00996	0.00268	0.00257
4	0.03686	0.03690	0.01019	0.01001	0.00270	0.00259
5	0.03728	0.03756	0.01031	0.01010	0.00273	0.00259
6	0.03774	0.03799	0.01045	0.01021	0.00277	0.00263
7	0.03821	0.03841	0.01058	0.01038	0.00281	0.00267
8	0.03869	0.03871	0.01072	0.01053	0.00284	0.00271

Table 3.3. Optimal bit number (quantization level) for the parameter vs length of the time series.

n	$s/n = 3$		$s/n = 4$		$s/n = 5$	
	m	v_ϕ	m	v_ϕ	m	v_ϕ
10	1	(2)	1	(2)	1	(2)
20	2	(4)	2	(3)	2	(3)
30	2	(4)	2	(4)	2	(4)
40	2	(5)	2	(5)	2	(5)
50	3	(6)	3	(6)	3	(6)
60	3	(6)	3	(6)	3	(6)
70	3	(7)	3	(7)	3	(7)
80	3	(7)	3	(7)	3	(7)
90	3	(8)	3	(7)	3	(7)
10	3	(8)	3	(8)	3	(8)

Symbols s, n and m are defined in Table 3.1, v_ϕ = quantization level for ϕ .

Table 3.4. Comparison of the mean square quantization errors from different schemes where the series consists several blocks with length $l \sim \text{Poisson}(\lambda)$, and parameter $\phi \sim U(0, 1 - 10^\delta)$, $\delta = 10^{-4}$.

	m	l	DPCM	Cummiskey's	Bayesian
$\lambda = 40$	3	30	0.06871	0.05920	0.03702
		40	0.07017	0.06658	0.03786
		50	0.07135	0.06567	0.03852
	4	30	0.01633	0.01841	0.01019
		40	0.01603	0.01642	0.01044
		50	0.01571	0.01335	0.01052
	5	30	0.00337	0.00637	0.00266
		40	0.00352	0.00669	0.00251
		50	0.00367	0.00476	0.00248
	3	70	0.07395	0.05110	0.03660
		80	0.07422	0.05122	0.03695
		90	0.07369	0.05004	0.03658
	4	70	0.01446	0.01343	0.00936
		80	0.01439	0.01344	0.00954
		90	0.01390	0.01319	0.00941
	5	70	0.00372	0.00430	0.00233
		80	0.00378	0.00422	0.00239
		90	0.00378	0.00485	0.00235

Table 3.5. Comparison of the mean square quantization errors where the series was generated by using initial parameter $\phi \sim U(0, 1 - \delta)$, $\delta = 10^{-4}$, and a random error $\Delta\phi \sim U(-0.1, 0.1)$ was added after each observation*.

m	l	DPCM	Cummiskey's	Bayesian
3	30	0.05385	0.05603	0.03039
	40	0.05390	0.06658	0.03081
	50	0.05421	0.06112	0.03110
	80	0.05485	0.04676	0.03175
	100	0.05530	0.04675	0.03175
4	30	0.01363	0.01570	0.00831
	40	0.01341	0.01471	0.00843
	50	0.01301	0.01301	0.00837
	80	0.01242	0.01234	0.00844
	100	0.01214	0.01244	0.00849
5	30	0.00308	0.00548	0.00225
	40	0.00316	0.00520	0.00223
	50	0.00320	0.00430	0.00222
	80	0.00329	0.00387	0.00222
	100	0.00333	0.00407	0.00224

* The new parameter $\phi + \Delta\phi$ is truncated by the upper and lower bounds 0 and $1 - \delta$ if it is out of range.

CHAPTER 4

BAYESIAN ADAPTIVE DPCM FOR THE TWO DIMENSIONAL AR(1,1) MODEL

4.1 Introduction

Two dimensional signal transmission, such as television broadcasting, facsimile transmission, tele-conference service and etc., has assumed a special role with increasing importance in our everyday life. So the research in this field has great practical significance. Quantization is an important step in two dimensional signal transmission and may affect the picture quality to a certain extent. Adaptive skill is also needed to deal with the nonstationary nature. The Bayesian adaptive DPCM scheme has been proved by simulation results to be a stable method in one dimensional signal quantization. It is desirable to extend the result to the two dimensional case.

A two dimensional AR(1,1) model is adopted to describe the data. The sample drawn from a picture is taken as a two dimensional sequence starting at its lower-left corner. The statistical property of the current observation is determined by the observed values one lag away in its neighborhood.

The derivation of the scheme and the optimal bit allocation is similar to the procedure in the one dimensional case. But due to the special characteristic of the two dimensional model, some adjustment is necessary for the extension.

A special AR(1,1) model adopted in our approach is introduced in 4.2 and some related features are discussed. In 4.3, the Bayesian adaptive DPCM scheme is extended to the two dimensional case. The simulation result is given in 4.4.

4.2 Two Dimensional AR(1,1) Model

The Bayesian adaptive DPCM can be extended to the two dimensional case for the stationary AR(1,1) model defined as

$$x_{ij} = \phi_1 x_{i-1,j} + \phi_2 x_{i,j-1} - \phi_1 \phi_2 x_{i-1,j-1} + \epsilon_{ij}, \quad (4.1)$$

where $|\phi_1| < 1, |\phi_2| < 1$ and $\{\epsilon_{ij}\}$ are i.i.d. normally distributed with mean zero and variance σ_ϵ^2 .

Model (4.1) has some nice properties, one of which is stated in the next theorem.

Theorem 4.2.1

Let $\{x_{ij}, \pm i = 0, 1, 2, \dots, \pm j = 0, 1, 2, \dots\}$ be a sample from model (4.1). Then,

$$x_{ij} = \phi_1^m x_{i-m,j} + \phi_2^n x_{i,j-n} - \phi_1^m \phi_2^n x_{i-m,j-n} + \epsilon'_{ij}, \quad (4.2)$$

where

$$\epsilon'_{ij} = \sum_{k=0}^{m-1} \sum_{l=0}^{n-1} \phi_1^k \phi_2^l \epsilon_{i-k,j-l}.$$

Proof of Theorem 4.2.1

This result can be proved by induction. Assume (4.2) holds for any integers m and n . We will prove it holds for $m+1$ and $n+1$.

The observations $x_{i-m,j}$ and $x_{i,j-n}$ can be iteratively expressed by the previous observations using model (4.1). We will see that two terms get canceled at each step

of the iteration.

$$\begin{aligned}
 & x_{i-m,j} \\
 = & \phi_1 x_{i-m-1,j} + \phi_2 x_{i-m,j-1} - \phi_1 \phi_2 x_{i-m-1,j-1} + \epsilon_{i-m,j} \\
 = & \phi_1 x_{i-m-1,j} + \phi_2^2 x_{i-m,j-2} - \phi_1 \phi_2^2 x_{i-m-1,j-2} + \epsilon_{i-m,j} + \phi_2 \epsilon_{i-m,j-1} \\
 & \dots \\
 = & \phi_1 x_{i-m-1,j} + \phi_2^n x_{i-m,j-n} - \phi_1 \phi_2^n x_{i-m-1,j-n} + \sum_{l=0}^{n-1} \phi_2^l \epsilon_{i-m,j-l} \quad (4.3)
 \end{aligned}$$

Similarly,

$$x_{i,j-n} = \phi_2 x_{i,j-n-1} + \phi_1^m x_{i-m,j-n} - \phi_1^m \phi_2 x_{i-m,j-n-1} + \sum_{k=0}^{m-1} \phi_1^k \epsilon_{i-k,j-n}. \quad (4.4)$$

Substituting (4.3) and (4.4) into (4.2), we have

$$\begin{aligned}
 x_{ij} &= \phi_1^m x_{i-m,j} + \phi_2^n x_{i,j-n} - \phi_1^m \phi_2^n x_{i-m,j-n} + \epsilon'_{ij} \\
 &= \phi_1^{m+1} x_{i-m-1,j} + \phi_1^m \phi_2^n x_{i-m,j-n} - \phi_1^{m+1} \phi_2^n x_{i-m-1,j-n} + \sum_{l=0}^{n-1} \phi_1^m \phi_2^l \epsilon_{i-m,j-l} \\
 &\quad + \phi_2^{n+1} x_{i,j-n-1} + \phi_1^m \phi_2^n x_{i-m,j-n} - \phi_1^m \phi_2^{n+1} x_{i-m,j-n-1} + \sum_{k=0}^{m-1} \phi_1^k \phi_2^n \epsilon_{i-k,j-n} \\
 &\quad - \phi_1^m \phi_2^n x_{i-m,j-n} + \sum_{k=0}^{m-1} \sum_{l=0}^{n-1} \phi_1^k \phi_2^l \epsilon_{i-k,j-l} \\
 &= \phi_1^{m+1} x_{i-m-1,j} + \phi_2^{n+1} x_{i,j-n-1} - \phi_1^{m+1} \phi_2^{n+1} x_{i-m-1,j-n-1} + \sum_{k=0}^m \sum_{l=0}^n \phi_1^k \phi_2^l \epsilon_{i-k,j-l}.
 \end{aligned}$$

Theorem 4.2.1 shows that the AR(1,1) model (4.1) holds no matter what lags m and n are used. Model (4.1) is a special case of (4.2) with $m = n = 1$.

For the one dimensional AR(1) model, an observation can be written as an infinite series of the previous noise. The variance and autocovariance can be easily computed

from the series expression. This result can be extended to the two dimensional model (4.1).

Lemma 4.2.1

Assume $\{x_{ij}, \pm i = 0, 1, 2, \dots, \pm j = 0, 1, 2, \dots\}$ are from the model (4.1). Then,

$$x_{ij} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \phi_1^k \phi_2^l \epsilon_{i-k, j-l}. \quad (4.5)$$

Proof of Lemma 4.2.1

By iterating model (4.1), we obtain

$$\begin{aligned} x_{ij} &= \phi_1 x_{i-1, j} + \phi_2 x_{i, j-1} - \phi_1 \phi_2 x_{i-1, j-1} + \epsilon_{ij} \\ &= \phi_1^2 x_{i-2, j} + \phi_1 \phi_2 x_{i-1, j-1} + \phi_2^2 x_{i, j-2} - \phi_1^2 \phi_2 x_{i-2, j-1} - \phi_1 \phi_2^2 x_{i-1, j-2} \\ &\quad + \phi_1 \epsilon_{i-1, j} + \phi_2 \epsilon_{i, j-1} + \epsilon_{ij} \\ &= \dots \\ &= \sum_{S_{n1}} \phi_1^k \phi_2^l x_{i-k, j-l} - \sum_{S_{n2}} \phi_1^k \phi_2^l x_{i-k, j-l} + \sum_{T_n} \phi_1^k \phi_2^l \epsilon_{i-k, j-l}, \end{aligned}$$

where T_n, S_{n1} and S_{n2} are sets of pairs of nonnegative integers and they are, with fixed integer n , defined by

$$\begin{aligned} T_n &= \{(k, l) : k \geq 0, l \geq 0, k + l < n\} \\ S_{n1} &= \{(k, l) : k \geq 0, l \geq 0, k + l = n\} \\ S_{n2} &= \{(k, l) : k > 0, l > 0, k + l = n + 1\}. \end{aligned}$$

Note that the sets S_{n1} and S_{n2} contain $n + 1$ and n elements, respectively. It is easy to check that

$$\begin{aligned}
& \sum_{S_{n1}} |\phi_1^k \phi_2^l| \\
&= \sum_{S_{n1}} |\phi_1^k \phi_2^{n-k}| \\
&\leq \sum_{S_{n1}} [\max(|\phi_1|, |\phi_2|)]^n \\
&= n \max(|\phi_1|^n, |\phi_2|^n) \\
&\rightarrow 0.
\end{aligned}$$

as $n \rightarrow \infty$ since $|\phi_1|^n < 1$ and $|\phi_2|^n < 1$. Similarly, $\sum_{S_{n2}} \phi_1^k \phi_2^l \rightarrow 0$.

By the inequality

$$(a_1 + a_2 + \dots + a_n)^2 \leq 2(a_1^2 + a_2^2 + \dots + a_n^2),$$

for any real $a_1, a_2, \dots, a_n$, we have

$$\begin{aligned}
& E \left[\left(x_{ij} - \sum_{T_n} \phi_1^k \phi_2^l \epsilon_{i-k, j-l} \right)^2 \right] \\
&= E \left[\left(\sum_{S_{n1}} \phi_1^k \phi_2^l x_{i-k, j-l} - \sum_{S_{n2}} \phi_1^k \phi_2^l x_{i-k, j-l} \right)^2 \right] \\
&\leq 2E \left[\sum_{S_{n1}} \phi_1^{2k} \phi_2^{2l} x_{i-k, j-l}^2 + \sum_{S_{n2}} \phi_1^{2k} \phi_2^{2l} x_{i-k, j-l}^2 \right]
\end{aligned}$$

$$\leq 2\sigma_x^2 \left[\sum_{S_{n1}} |\phi_1^{2k} \phi_2^{2l}| \sum_{S_{n2}} |\phi_1^{2k} \phi_2^{2l}| \right]$$

$$\rightarrow 0,$$

as $n \rightarrow \infty$ by the previous result.

Hence, $\sum_{T_n} \phi_1^k \phi_2^l \epsilon_{i-k, j-l}$ converges to x_{ij} in mean square by the Cauchy criterion. We conclude that

$$x_{ij} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \phi_1^k \phi_2^l \epsilon_{i-k, j-l}.$$

The next theorem gives the variance and autocovariance of $\{x_{ij}\}$ by applying its series expansion.

Theorem 4.2.2

Assume $\{x_{ij}, \pm i = 0, 1, 2, \dots, \pm j = 0, 1, 2, \dots\}$ are from the model (4.1). Then,

$$\sigma_x^2 = \text{Var}(x_{ij}) = \frac{\sigma_\epsilon^2}{(1 - \phi_1^2)(1 - \phi_2^2)}$$

$$\gamma(p, q) = E(x_{ij} x_{i+p, j+q}) = \phi_1^{|p|} \phi_2^{|q|} \sigma_x^2. \quad (4.6)$$

Proof of Theorem 4.2.2

By lemma 4.2.1,

$$\sigma_x^2 = E \left[\left(\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \phi_1^k \phi_2^l \epsilon_{i-k, j-l} \right)^2 \right]$$

$$= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \phi_1^{2k} \phi_2^{2l} E \left[(\epsilon_{i-k, j-l})^2 \right]$$

$$= \frac{\sigma_\epsilon^2}{(1 - \phi_1^2)(1 - \phi_2^2)}$$

since the ϵ_{ij} 's are all independent.

For positive p and q ,

$$\begin{aligned} \gamma(p, q) &= E(x_{ij}x_{i-p,j-q}) \\ &= E \left[\left(\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \phi_1^k \phi_2^l \epsilon_{i-k,j-l} \right) \left(\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \phi_1^k \phi_2^l \epsilon_{i-p-k,j-q-l} \right) \right] \\ &= E \left[\left(\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \phi_1^k \phi_2^l \epsilon_{i-k,j-l} \right) \left(\sum_{k=p}^{\infty} \sum_{l=q}^{\infty} \phi_1^{k-p} \phi_2^{l-q} \epsilon_{i-k,j-l} \right) \right] \\ &= \sum_{k=p}^{\infty} \sum_{l=q}^{\infty} \phi_1^{2k-p} \phi_2^{2l-q} \sigma_\epsilon^2 \\ &= \phi_1^{-p} \phi_2^{-q} \frac{\phi_1^{2p} \phi_2^{2q}}{(1 - \phi_1^2)(1 - \phi_2^2)} \sigma_\epsilon^2 \\ &= \phi_1^p \phi_2^q \sigma_x^2. \end{aligned}$$

So

$$\gamma(p, q) = \phi_1^{|p|} \phi_2^{|q|} \sigma_x^2$$

by symmetry.

Multiplying $x_{i-1,j}$ and $x_{i,j-1}$ respectively to both sides of (4.1) and then taking the expectation, we get the Yule-Walker equation

$$\gamma(1, 0) = \phi_1 \gamma(0, 0) + \phi_2 \gamma(1, 1) - \phi_1 \phi_2 \gamma(0, 1)$$

$$\gamma(0,1) = \phi_1\gamma(1,1) + \phi_2\gamma(0,0) - \phi_1\phi_2\gamma(1,0). \quad (4.7)$$

for this two dimensional AR(1,1) model. Estimates of the parameters $\hat{\phi}_1$ and $\hat{\phi}_2$ can be obtained by substituting the sample variance and autocovariance into (4.7) and solving it with respect to ϕ_1 and ϕ_2 .

The least square predictor for x_{ij} based on the previous observations is

$$x_{ij} = \phi_1 x_{i-1,j} + \phi_2 x_{i,j-1} - \phi_1 \phi_2 x_{i-1,j-1} \quad (4.8)$$

with mean squared prediction error σ_e^2 .

4.3 Two Dimensional Bayesian Adaptive DPCM

The derivation of the Bayesian adaptive DPCM for the two dimensional AR(1,1) model is similar to that of the one dimensional AR(1) model. So similar notation is defined and used without further explanation. Some results can be derived in a similar way and they are just stated without proof. Due to the difference between the two models, the derivation is not just a simple repetition of the same procedure. Detailed proof is given in case it is needed.

Let $\{x_{ij}, i = 1, 2, \dots, n_1, j = 1, 2, \dots, n_2\}$ be a sample from model (4.1). Considering that the sample we are dealing with is finite, we rewrite the truncated version of model (4.1) in vector form

$$\begin{cases} x_{ij} \sim N(0, \sigma_x^2), & \text{if } i = 1 \text{ or } j = 1 \\ x_{ij} = \Phi' X_{ij} + \epsilon_{ij}, & \text{otherwise,} \end{cases} \quad (4.9)$$

where $\Phi' = (\phi_1, \phi_2, -\phi_1\phi_2)$ and $X_{ij}' = (x_{i-1,j}, x_{i,j-1}, x_{i-1,j-1})$. We assume ϕ_1, ϕ_2 and the white noise are all independent and assign a joint prior density function $\pi_k(\phi_1, \phi_2)$ to ϕ_1 and ϕ_2 .

Define $\epsilon_{ij} = x_{ij}$, for $i = 1$ or $j = 1$. The prediction residual is given by

$$e_{ij} = \begin{cases} x_{ij}, & \text{if } i = 1 \text{ or } j = 1 \\ x_{ij} - \hat{\Phi}'X_{ij}, & \text{otherwise.} \end{cases} \quad (4.10)$$

Since only quantized values $\{\hat{e}_{ij}\}$, $\hat{\phi}_1$ and $\hat{\phi}_2$ are available on the receiver side, the reconstructed value for x_{ij} is

$$\hat{x}_{ij} = \hat{\Phi}'\hat{X}_{ij} + \hat{e}_{ij}. \quad (4.11)$$

It is the direct result of equations (4.10) and (4.11) that the reconstruction error is equal to the quantization error of the residual x_{ij} , i.e.

$$e_{ij} - \hat{e}_{ij} = x_{ij} - \hat{x}_{ij}. \quad (4.12)$$

Just as in the one dimensional case, the prediction residuals $\{e_{ij}\}$ tend to the white noise ϵ_{ij} in mean square as the quantization levels tend to infinity and they are asymptotically independent.

We can see that model (4.9) has a similar form to the one dimensional AR(p) model (3.2). But there are indeed some differences. One important feature of the two dimensional model is that the vector Φ has three components but only two of them can be determined independently. We need the next lemma to show the quantized vector $\hat{\Phi}$ is orthogonal to the quantization error $\Phi - \hat{\Phi}$.

Lemma 4.3.1

Assume ϕ_1 and ϕ_2 are the parameters for model (4.9) and $\hat{\phi}_1$ and $\hat{\phi}_2$ are the L - M quantizer. Then,

$$E[(\phi_1\phi_2 - \hat{\phi}_1\hat{\phi}_2)\hat{\phi}_1\hat{\phi}_2] = 0. \quad (4.13)$$

Proof of Lemma 4.3.1

$$\begin{aligned}
& E[(\phi_1\phi_2 - \hat{\phi}_1\hat{\phi}_2)\hat{\phi}_1\hat{\phi}_2] \\
&= E[(\phi_1\phi_2 - \hat{\phi}_1\phi_2 + \hat{\phi}_1\phi_2 - \hat{\phi}_1\hat{\phi}_2)\hat{\phi}_1\hat{\phi}_2] \\
&= E[(\phi_1\phi_2 - \hat{\phi}_1\phi_2)\hat{\phi}_1\hat{\phi}_2] + E[(\hat{\phi}_1\phi_2 - \hat{\phi}_1\hat{\phi}_2)\hat{\phi}_1\hat{\phi}_2] \\
&= E[(\phi_1 - \hat{\phi}_1)\hat{\phi}_1\phi_2\hat{\phi}_2] + E[(\phi_2 - \hat{\phi}_2)\hat{\phi}_1^2\hat{\phi}_2] \\
&= 0
\end{aligned}$$

by Lemma 3.2.2.

The next theorem proves the asymptotic independence between $\hat{X}_{ij}$ and $(X_{ij} - \hat{X}_{ij})$ for fixed ϕ_1 and ϕ_2 .

Theorem 4.3.1

Let X_{ij} be defined as in model (4.9) and $\hat{X}_{ij}$ is the L - M quantizer for X_{ij} . Then,

$$E \left[\hat{X}_{ij} (X_{ij} - \hat{X}_{ij})' | \Phi \right] \rightarrow 0, \quad (4.14)$$

as the quantization levels tend to infinity.

Proof of Theorem 4.3.1

$$\begin{aligned}
& \hat{X}_{ij} (X_{ij} - \hat{X}_{ij})' \\
&= \begin{pmatrix} (x_{i-1,j} - \hat{x}_{i-1,j})\hat{x}_{i-1,j} & (x_{i-1,j} - \hat{x}_{i-1,j})\hat{x}_{i,j-1} & (x_{i-1,j} - \hat{x}_{i-1,j})\hat{x}_{i-1,j-1} \\ (x_{i,j-1} - \hat{x}_{i,j-1})\hat{x}_{i,j-1} & (x_{i,j-1} - \hat{x}_{i,j-1})\hat{x}_{i,j-1} & (x_{i,j-1} - \hat{x}_{i,j-1})\hat{x}_{i-1,j-1} \\ (x_{i-1,j-1} - \hat{x}_{i-1,j-1})\hat{x}_{i-1,j} & (x_{i-1,j-1} - \hat{x}_{i-1,j-1})\hat{x}_{i,j-1} & (x_{i-1,j-1} - \hat{x}_{i-1,j-1})\hat{x}_{i-1,j-1} \end{pmatrix}.
\end{aligned}$$

All the components of the matrix have the form of $(x_{ij} - \hat{x}_{ij})\hat{x}_{kl}$, and they may be divided into three groups. We now prove that, in each group, the conditional expectations tend to zero.

If $i = k$ and $j = l$, the conditional expectation vanishes by the property of the L - M quantizer.

If $i = k + 1$ or $j = l + 1$,

$$(x_{ij} - \hat{x}_{ij}) = (e_{ij} - \hat{e}_{ij}) \rightarrow (\epsilon_{ij} - \hat{e}_{ij}),$$

which is independent of x_{kl} . So the result holds.

If $i = k - 1$ or $j = l - 1$, the $\hat{x}_{kl}$ can be written as a linear combination of the prediction errors and $\{\hat{x}_{pq}, p < i \text{ or } q < j\}$. All terms in the combination are asymptotically independent of $(x_{ij} - \hat{x}_{ij})$ except the one containing $\hat{e}_{ij}$. The conditional expectation of this term vanishes by the result in case of $i = k$ and $j = l$.

This completes the proof.

We rewrite e_{ij} as

$$\begin{aligned} e_{ij} &= x_{ij} - \hat{\Phi}'\hat{X}_{ij} \\ &= \epsilon_{ij} + (\Phi - \hat{\Phi})'\hat{X}_{ij} + \Phi'(X_{ij} - \hat{X}_{ij}). \end{aligned}$$

by Theorem 4.3.1 and the fact that $\{\epsilon_{ij}\}$ is independent of the parameters and the previous observations. Then,

$$\sigma_e^2 = E(e_{ij}^2)$$

$$\begin{aligned}
&= \sigma_e^2 + E \left[(\Phi - \hat{\Phi})' \hat{X}_{ij} \hat{X}_{ij}' (\Phi - \hat{\Phi}) \right] + E \left[\Phi' (X_{ij} - \hat{X}_{ij}) (X_{ij} - \hat{X}_{ij})' \Phi \right] \\
&= \sigma_e^2 + E \left[(\Phi - \hat{\Phi})' X_{ij} X_{ij}' (\Phi - \hat{\Phi}) \right] + E \left[\hat{\Phi}' (X_{ij} - \hat{X}_{ij}) (X_{ij} - \hat{X}_{ij})' \hat{\Phi} \right]
\end{aligned}$$

by Lemma 4.3.1 and Theorem 4.3.1.

Suppose that we have totally s bits available and m_k bits will be assigned to the parameter ϕ_k . So the quantization level for ϕ_k is $v_{\phi_k} = 2^{m_k}$. For the b bits left after assigning m bits to the parameters, let b_1 and b_2 be the numbers of bits assigned to the observations x_{ij} with index $i = 1$ or $j = 1$ and the prediction residuals respectively. Then, by Huang and Schultheiss' optimal bit allocation, b_1 and b_2 can be calculated by

$$\begin{aligned}
b_1 &= \frac{b}{n_1 n_2} + \frac{1}{2} \log_2 [E(\gamma_0)]^{\frac{n_1 n_2 - n_1 - n_2 + 1}{n_1 n_2}} \\
b_2 &= \frac{b}{n_1 n_2} + \frac{1}{2} \log_2 [E(\gamma_0)]^{-\frac{n_1 + n_2 - 1}{n_1 n_2}}
\end{aligned} \tag{4.15}$$

where $\gamma_0 = \sigma_x^2 / \sigma_e^2$, and the corresponding quantization levels are

$$\begin{aligned}
v_1 &= 2^{b_1} = 2^{\frac{b}{n_1 n_2}} [E(\gamma_0)]^{\frac{n_1 n_2 - n_1 - n_2 + 1}{2 n_1 n_2}} \\
v_2 &= 2^{b_2} = 2^{\frac{b}{n_1 n_2}} [E(\gamma_0)]^{-\frac{n_1 + n_2 - 1}{2 n_1 n_2}},
\end{aligned} \tag{4.16}$$

respectively.

We now write σ_e^2 as a function of the quantization levels $\{v_{\phi_k}\}$, v_1 and v_2 by using the next two lemmas.

Lemma 4.3.2

Suppose $\{x_{ij}, i = 1, 2, \dots, n_1, j = 1, 2, \dots, n_2\}$ is a sample from the model (4.9).

Then,

$$E[(\Phi - \hat{\Phi})' X_{ij} X_{ij}' (\Phi - \hat{\Phi})] \simeq \sigma_e^2 \sum_{k=1}^2 \frac{G_k}{v_{\phi_k}^2}, \quad (4.17)$$

where

$$G_k = \frac{1}{12} \left[\int_{-\infty}^{\infty} \pi_k^{\frac{1}{2}}(\phi_k) d\phi_k \right]^2 \left[\int_{-\infty}^{\infty} \frac{1}{(1 - \phi_k^2)} \pi_k^{\frac{1}{2}}(\phi_k) d\phi_k \right].$$

Proof of Lemma 4.3.2

Let P be the correlation matrix of X_{ij} . Expanding the product $(\Phi - \hat{\Phi})' P (\Phi - \hat{\Phi})$, we have

$$\begin{aligned} & (\Phi - \hat{\Phi})' P (\Phi - \hat{\Phi}) \\ &= \begin{pmatrix} (\phi_1 - \hat{\phi}_1) \\ (\phi_2 - \hat{\phi}_2) \\ -(\phi_1 \phi_2 - \hat{\phi}_1 \hat{\phi}_2) \end{pmatrix}' \begin{pmatrix} 1 & \phi_1 \phi_2 & \phi_2 \\ \phi_1 \phi_2 & 1 & \phi_1 \\ \phi_2 & \phi_1 & 1 \end{pmatrix} \begin{pmatrix} (\phi_1 - \hat{\phi}_1) \\ (\phi_2 - \hat{\phi}_2) \\ -(\phi_1 \phi_2 - \hat{\phi}_1 \hat{\phi}_2) \end{pmatrix} \\ &= (\phi_1 - \hat{\phi}_1)^2 + (\phi_2 - \hat{\phi}_2)^2 - \phi_1^2 \phi_2^2 + \hat{\phi}_1^2 \hat{\phi}_2^2 \\ &\quad + 2\phi_1 \hat{\phi}_1 \hat{\phi}_2 (\phi_2 - \hat{\phi}_2) + 2\phi_2 \hat{\phi}_1 \hat{\phi}_2 (\phi_1 - \hat{\phi}_1). \end{aligned}$$

So

$$E[(\Phi - \hat{\Phi})' X_{ij} X_{ij}' (\Phi - \hat{\Phi})]$$

$$\begin{aligned}
&= \sigma_\epsilon^2 E \left[\frac{(\phi_1 - \hat{\phi}_1)^2 + (\phi_2 - \hat{\phi}_2)^2 - \phi_1^2 \phi_2^2 + \hat{\phi}_1^2 \hat{\phi}_2^2}{(1 - \phi_1^2)(1 - \phi_2^2)} \right] \\
&\quad + 2\sigma_\epsilon^2 E \left[\frac{\phi_1 \hat{\phi}_1 \hat{\phi}_2 (\phi_2 - \hat{\phi}_2) + \phi_2 \hat{\phi}_1 \hat{\phi}_2 (\phi_1 - \hat{\phi}_1)}{(1 - \phi_1^2)(1 - \phi_2^2)} \right].
\end{aligned}$$

The second term tends to zero by Lemma 3.5.2. We can consider only the first term.

Note that

$$E \left[\frac{\phi_k^2 - \hat{\phi}_k^2}{1 - \phi_k^2} \right] \simeq E \left[\frac{(\phi_k - \hat{\phi}_k)^2}{1 - \phi_k^2} \right],$$

for $k = 1, 2$ when v_{ϕ_1} and v_{ϕ_2} are large. By adding and subtracting $\hat{\phi}_1^2 \hat{\phi}_2^2$ in the numerator,

$$\begin{aligned}
&E \left[\frac{(\phi_1 - \hat{\phi}_1)^2 + (\phi_2 - \hat{\phi}_2)^2 - \phi_1^2 \phi_2^2 + \hat{\phi}_1^2 \hat{\phi}_2^2}{(1 - \phi_1^2)(1 - \phi_2^2)} \right] \\
&= E \left[\frac{(\phi_1 - \hat{\phi}_1)^2 + (\phi_2 - \hat{\phi}_2)^2 - (\phi_1^2 - \hat{\phi}_1^2) \phi_2^2 - \hat{\phi}_1^2 (\phi_2^2 - \hat{\phi}_2^2)}{(1 - \phi_1^2)(1 - \phi_2^2)} \right] \\
&\simeq E \left[\frac{(\phi_1 - \hat{\phi}_1)^2 (1 - \phi_2^2) + (\phi_2 - \hat{\phi}_2)^2 (1 - \hat{\phi}_2^2)}{(1 - \phi_1^2)(1 - \phi_2^2)} \right] \\
&\simeq E \left[\frac{(\phi_1 - \hat{\phi}_1)^2}{(1 - \phi_1^2)} \right] + E \left[\frac{(\phi_2 - \hat{\phi}_2)^2}{(1 - \phi_2^2)} \right] \\
&= \sum_{k=1}^2 \frac{G_k}{v_{\phi_k}^2},
\end{aligned}$$

by Lemma 3.5.1.

Lemma 4.3.3

Suppose $\{x_{ij}, i = 1, 2, \dots, n_1, j = 1, 2, \dots, n_2\}$ is a sample from model (4.9).

Then,

$$\begin{aligned} & E \left[\hat{\Phi}'(X_{ij} - \hat{X}_{ij})(X_{ij} - \hat{X}_{ij})' \hat{\Phi} \right] \\ &= \left\{ \sum_{k=1}^2 \left[E(\phi_k^2) - \frac{D_k}{v_{\phi_k}^2} \right] + \prod_{k=1}^2 \left[E(\phi_k^2) - \frac{D_k}{v_{\phi_k}^2} \right] \right\} I \sigma_\epsilon^2 v_2^{-2}. \end{aligned}$$

Proof of Lemma 4.3.3

Note that the reconstruction error is equal to the quantization error which is independent of the parameters. The mean of the cross product terms tend to zero as the quantization levels go to infinity. So

$$\begin{aligned} & E \left[\hat{\Phi}'(X_{ij} - \hat{X}_{ij})(X_{ij} - \hat{X}_{ij})' \hat{\Phi} \right] \\ &= E \left[\hat{\phi}_1^2(x_{i-1,j} - \hat{x}_{i-1,j})^2 + \hat{\phi}_2^2(x_{i,j-1} - \hat{x}_{i,j-1})^2 + \hat{\phi}_1^2 \hat{\phi}_1^2(x_{i-1,j-1} - \hat{x}_{i-1,j-1})^2 \right] \\ &= I v_2^{-2} \sigma_\epsilon^2 [E(\hat{\phi}_1^2) + E(\hat{\phi}_2^2) + E(\hat{\phi}_1^2)E(\hat{\phi}_2^2)] \\ &= \left\{ \sum_{k=1}^2 \left[E(\phi_k^2) - \frac{D_k}{v_{\phi_k}^2} \right] + \prod_{k=1}^2 \left[E(\phi_k^2) - \frac{D_k}{v_{\phi_k}^2} \right] \right\} I \sigma_\epsilon^2 v_2^{-2}. \end{aligned}$$

From Lemma 4.3.2 and Lemma 4.3.3, we have

$$\sigma_\epsilon^2 = C(v_{\phi_1}, v_{\phi_2}, v_1, v_2) \sigma_\epsilon^2, \quad (4.18)$$

where

$$C(v_{\phi_1}, v_{\phi_2}, v_1, v_2) = \frac{1 + \sum_{k=1}^2 G_k v_{\phi_k}^{-2}}{1 - \left\{ \sum_{k=1}^2 \left[E(\phi_k^2) - D_k v_{\phi_k}^{-2} \right] + \prod_{k=1}^2 \left[E(\phi_k^2) - D_k v_{\phi_k}^{-2} \right] \right\} I v_2^{-2}}. \quad (4.19)$$

So by Lemma 3.2.1, the mean squared quantization error has the expression

$$E[(e_{ij} - \hat{e}_{ij})^2] = \sigma_\epsilon^2 \frac{1 + \sum_{k=1}^2 G_k v_{\phi_k}^{-2}}{v_2^2 I^{-1} - \sum_{k=1}^2 \left[E(\phi_k^2) - D_k v_{\phi_k}^{-2} \right] - \prod_{k=1}^2 \left[E(\phi_k^2) - D_k v_{\phi_k}^{-2} \right]}. \quad (4.20)$$

By Theorem 3.4.1, the total mean square quantization error is

$$\begin{aligned} & E\left[\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (e_{ij} - \hat{e}_{ij})^2\right] \\ &= n_1 n_2 \sigma_\epsilon^2 \frac{1 + \sum_{k=1}^2 G_k v_{\phi_k}^{-2}}{v_2^2 I^{-1} - \sum_{k=1}^2 \left[E(\phi_k^2) - D_k v_{\phi_k}^{-2} \right] - \prod_{k=1}^2 \left[E(\phi_k^2) - D_k v_{\phi_k}^{-2} \right]} \\ &= n_1 n_2 \sigma_\epsilon^2 H, \end{aligned}$$

say.

The best choice of $\{v_{\phi_k}\}$ can be obtained by differentiating H with respect to $\{v_{\phi_k}\}$ and then setting the derivative to be zero. Note that

$$v_{\phi_k} = 2^{m_k},$$

$$v_2 = 2^{\frac{s - \sum_{k=1}^2 m_k}{\frac{k=1}{n_1 n_2}}} [E(\gamma_0)]^{-\frac{p}{2n_1 n_2}}$$

$$= v_{\phi_h}^{-\frac{1}{n_1 n_2}} 2^{\frac{s - \sum_{k \neq j}^2 m_j}{n_1 n_2}} [E(\gamma_0)]^{-\frac{n_1 + n_2 - 1}{2 n_1 n_2}}.$$

So

$$\frac{\partial v_2^2}{\partial v_{\phi_h}^2} = -\frac{v_2^2}{n_1 n_2 v_{\phi_h}^2}.$$

The optimal $\{v_{\phi_h}\}$ will be the solution for

$$\frac{\partial U}{\partial v_{\phi_h}} W - U \frac{\partial W}{\partial v_{\phi_h}} = 0, \quad (4.21)$$

for $k = 1, 2$ subject to $\frac{v_1}{v_2} = [E(\gamma_0)]^{\frac{1}{2}}$ and $m + b = s$, where

$$U = 1 + \sum_{k=1}^2 G_k v_{\phi_h}^{-2}$$

$$W = v_2^2 I^{-1} - \sum_{k=1}^2 [E(\phi_k^2) - D_k v_{\phi_h}^{-2}] - \prod_{k=1}^2 [E(\phi_k^2) - D_k v_{\phi_h}^{-2}].$$

4.4 Simulation

The approximation for the mean square quantization error using (4.9) and the simulation result are listed in Table 4.1. In the simulation, the parameters ϕ_1 and ϕ_2 were generated uniformly on the interval $(0, 1 - \delta]$ with $\delta = 10^{-2}$. Data sets with $n_1 = n_2 = 16, 32, 48$, and 64 were generated and the white noise is normally distributed with zero mean and unit variance. Since $n_1 = n_2$ and ϕ_1 and ϕ_2 have same distribution, we assign the same number of bits for quantizing ϕ_1 and ϕ_2 . The simulation was run 1000 times with average bits $s/(n_1 n_2) = 3, 4$, and 5.

We can see that formula (4.20) gives good approximation for the quantization error even if the average bit number for each observation is not so large. The difference between the approximation and the simulation gets smaller as the bit number increases. The optimal number of bits assigned to the parameters determined by (4.20) coincides with or is very close to that from simulation. So the approximation formula can be used to determine the bit allocation.

Table 4.1. Mean squared quantization errors from asymptotic approximation $(E[(e_{ij} - \hat{e}_{ij})^2])$ and simulation $(\hat{E}[(e_{ij} - \hat{e}_{ij})^2])$. $\phi_1, \phi_2 \sim U(0, 1 - \delta)$, $\delta = 10^{-2}$.

$n_1 = n_2 = 16$

m_k	$s/(n_1 n_2) = 3$		$s/(n_1 n_2) = 4$		$s/(n_1 n_2) = 5$	
	$E[(e_{ij} - \hat{e}_{ij})^2]$	$\hat{E}[(e_{ij} - \hat{e}_{ij})^2]$	$E[(e_{ij} - \hat{e}_{ij})^2]$	$\hat{E}[(e_{ij} - \hat{e}_{ij})^2]$	$E[(e_{ij} - \hat{e}_{ij})^2]$	$\hat{E}[(e_{ij} - \hat{e}_{ij})^2]$
0	0.06125	0.13180	0.01703	0.03466	0.00453	0.00842
1	0.04799	0.06947	0.01328	0.01743	0.00353	0.00458
2	0.04494	0.05035	0.01242	0.01347	0.00330	0.00373
3	0.04449	0.04535	0.01230	0.01237	0.00327	0.00333
4	0.04471	0.04450	0.01236	0.01228	0.00329	0.00329
5	0.04510	0.04477	0.01247	0.01236	0.00332	0.00331
6	0.04553	0.04514	0.01260	0.01248	0.00335	0.00334
7	0.04598	0.04561	0.01272	0.01261	0.00338	0.00337
8	0.04644	0.04599	0.01285	0.01273	0.00342	0.00340
9	0.04690	0.04641	0.01299	0.01285	0.00346	0.00343
10	0.04737	0.04692	0.01312	0.01297	0.00349	0.00346

s = total number of bits available, m_k = number of bits assigned to ϕ_k , n_1, n_2 = length and width of block.

Table 4.1. Continued. $n_1 = n_2 = 32$

m_k	$s(n_1, n_2)=3$		$s(n_1, n_2)=4$		$s(n_1, n_2)=5$	
	$E[(e_{ij}-e_{ij}^*)^2]$	$\hat{E}[(e_{ij}-e_{ij}^*)^2]$	$E[(e_{ij}-e_{ij}^*)^2]$	$E[(e_{ij}-e_{ij}^*)^2]$	$E[(e_{ij}-e_{ij}^*)^2]$	$E[(e_{ij}-e_{ij}^*)^2]$
0	0.05514	0.10480	0.01527	0.02579	0.00405	0.00734
1	0.04285	0.05915	0.01181	0.01462	0.00313	0.00381
2	0.03982	0.04376	0.01096	0.01155	0.00290	0.00332
3	0.03913	0.03999	0.01077	0.01098	0.00285	0.00299
4	0.03903	0.03939	0.01075	0.01090	0.00284	0.00292
5	0.03908	0.03940	0.01076	0.01089	0.00285	0.00294
6	0.03916	0.03949	0.01078	0.01090	0.00286	0.00294
7	0.03926	0.03959	0.01081	0.01094	0.00286	0.00294
8	0.03935	0.03966	0.01084	0.01096	0.00287	0.00295
9	0.03945	0.03979	0.01087	0.01099	0.00288	0.00295
10	0.03955	0.03987	0.01089	0.01103	0.00289	0.00296

Table 4.1. Continued. $n_1 = n_2 = 48$

m_k	$s/(n_1 n_2)=3$		$s/(n_1 n_2)=4$		$s/(n_1 n_2)=5$	
	$E[(e_{ij}-e_{ij}^*)^2]$	$\hat{E}[(e_{ij}-e_{ij}^*)^2]$	$E[(e_{ij}-e_{ij}^*)^2]$	$\hat{E}[(e_{ij}-e_{ij}^*)^2]$	$E[(e_{ij}-e_{ij}^*)^2]$	$\hat{E}[(e_{ij}-e_{ij}^*)^2]$
0	0.05320	0.10935	0.01471	0.03116	0.00390	0.00544
1	0.04127	0.06611	0.01136	0.01584	0.00301	0.00361
2	0.03830	0.04659	0.01053	0.01175	0.00279	0.00292
3	0.03758	0.03949	0.01033	0.01082	0.00273	0.00277
4	0.03744	0.03801	0.01029	0.01063	0.00272	0.00275
5	0.03743	0.03794	0.01029	0.01058	0.00272	0.00275
6	0.03746	0.03799	0.01030	0.01059	0.00272	0.00275
7	0.03750	0.03801	0.01031	0.01059	0.00273	0.00275
8	0.03754	0.03803	0.01032	0.01059	0.00273	0.00275
9	0.03758	0.03811	0.01033	0.01061	0.00273	0.00276
10	0.03762	0.03814	0.01035	0.01061	0.00274	0.00276

Table 4.1. Continued. $n_1 = n_2 = 64$

m_k	$s(n, n_2)=3$		$s(n, n_2)=4$		$s(n, n_2)=5$	
	$E[(e_{ij}-e_{ij}')^2]$	$\hat{E}[(e_{ij}-e_{ij}')^2]$	$E[(e_{ij}-e_{ij}')^2]$	$\hat{E}[(e_{ij}-e_{ij}')^2]$	$E[(e_{ij}-e_{ij}')^2]$	$\hat{E}[(e_{ij}-e_{ij}')^2]$
0	0.05224	0.10946	0.01444	0.02266	0.00382	0.00537
1	0.04051	0.06233	0.01115	0.01363	0.00295	0.00356
2	0.03757	0.04475	0.01033	0.01080	0.00273	0.00285
3	0.03685	0.03945	0.01013	0.01024	0.00268	0.00270
4	0.03669	0.03766	0.01008	0.01017	0.00266	0.00268
5	0.03666	0.03748	0.01007	0.01016	0.00266	0.00268
6	0.03668	0.03751	0.01008	0.01015	0.00266	0.00268
7	0.03670	0.03752	0.01008	0.01016	0.00267	0.00268
8	0.03672	0.03754	0.01009	0.01017	0.00267	0.00268
9	0.03674	0.03755	0.01010	0.01017	0.00267	0.00268
10	0.03676	0.03755	0.01010	0.01019	0.00267	0.00269

REFERENCES

- [1] R. G. Gallager. *Information theory and reliable communication*. New York, Wiley, 1968.
- [2] A. K. Jain. Image data compression: a review. *Proc. IEEE*, vol. 69, no. 3, pp. 349-389, Mar. 1981.
- [3] A. N. Netravali and J. O. Limb. Picture coding: a review. *Proc. IEEE*, vol. 68, no. 3, pp. 366-406, Mar. 1980.
- [4] J. R. Pierce B. M. Oliver and C. E. Shannon. The philosophy of PCM. *Proc. IRE*, vol. 36, pp. 1324-1331, Nov. 1948.
- [5] J. Max. Quantizing for minimum distortion. *IRE Trans. Inform. Theory*, vol. IT-5, pp. 7-12, Mar. 1960.
- [6] T. J. Goblick Jr. and J. L. Hoslinger. Analog source digitization: a comparison of theory and practice. *IEEE Trans. Inform. Theory (Corresp)*, vol. IT-13, pp. 323-326, Apr. 1967.
- [7] L. Zetterberg. A comparison between delta and pulse code modulation. *Ericsson Tech.*, vol. 2, no. 1, 1955.
- [8] N. C. Jayant. Digital coding of speech waveforms: PCM, DPCM and DM quantizers. *Proc. IEEE*, vol. 62, pp. 611-632, May 1974.
- [9] B. N. Oliver. Efficient coding. *Bell Syst. Tech. J.*, vol. 31, pp. 724-750, July 1952.
- [10] P. Elias. Predictive coding. *IRE Trans. Inform. Theory*, vol. IT-1, pp. 16-33, Mar. 1955.
- [11] J. B. O'Neal Jr. Predictive quantizing systems (differential pulse code modulation) for the transmission of television signals. *Bell Syst. Tech. J.*, vol. 45, pp. 689-721, May-June 1966.
- [12] P. A. Wintz. Transform picture coding. *Proc. IEEE*, vol. 60, pp. 809-820, July 1972.
- [13] S. J. Campanella and G. S. Robinson. A comparison of orthogonal transformations for digital speech processing. *IEEE Trans. Commun. Tech.*, vol. COM-19, pp. 1045-1049, Dec. 1971.

- [14] W. H. Chen and C. H. Smith. Adaptive coding of monochrome and color images. *IEEE Trans. Commun. Tech.*, vol. COM-25, no.11, pp. 1285-1292, Nov. 1977.
- [15] J. J. Y. Huang and P. M. Schultheiss. Block quantization of correlated Gaussian random variables. *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 280-296, Sept. 1963.
- [16] N. S. Jayant. Adaptive quantization with one-word memory. *Bell Syst. Tech. J.*, vol. 52, pp. 1119-1144, Sept. 1973.
- [17] C. S. Golding and P. M. Schultheiss. Study of an adaptive quantizer. *Proc. IEEE*, vol. 55, pp. 293-297, Mar. 1967.
- [18] P. Cummiskey. *Adaptive differential pulse-code modulation for speech processing*. PhD thesis, Newark College of Engineering, Newark, N.J., 1973.
- [19] B. S. Atal and M. R. Schroeder. Adaptive predictive coding of speech signals. *Bell Syst. Tech. J.*, vol. 49, pp. 1973-1986, Oct. 1970.
- [20] R. W. Stroh. *Optimum and adaptive differential PCM*. PhD thesis, Polytech. Inst. Brooklyn, Farmingdale, N.Y., 1970.
- [21] J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer-Verlag, New York, 2nd edition, 1985.
- [22] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, vol. 27, pp. 379-423, 623-656, 1970.
- [23] S. P. Lloyd. Least squares quantization in PCM. *IEEE Trans. Inform. Theory*, vol. IT-28, no. 2, pp. 129-137, Mar. 1982.
- [24] J. A. Bucklew and Jr. N. C. Gallagher. A note on optimal quantization. *IEEE Trans. Inform. Theory*, vol. IT-25, no. 3, pp. 365-366, May 1979.
- [25] M. D. Paez and T. H. Glisson. Minimum mean-squared-error quantization in speech PCM and DPCM systems. *IEEE Trans. Commun. Tech.*, vol. COM-20, pp. 225-230, Apr. 1972.
- [26] H. L. Royden. *Real analysis*. Macmillan, New York, 2nd edition, 1963.
- [27] G. M. Roe. Quantizing for minimum distortion. *IEEE Trans. Inform. Theory*, vol. IT-10, pp. 384-385, Oct. 1964.
- [28] P. F. Panter and W. Dite. Quantizing distortion in pulse-code modulation with nonuniform spacing levels. *Proc. IRE*, vol. 39, pp. 44-48, 1951.
- [29] C. C. Cutler. Differential quantization of communications. U.S. Patent 2 605 361, July 29 1972.
- [30] L. Goldstein and Bede Liu. Quantization error and step-size distribution in ADPCM. *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 216-223, Mar. 1977.

- [31] D. S. Arnstein. Quantization error in predictive coders. *IEEE Trans. Commun.*, vol. COM-23, pp. 423-429, Apr. 1975.
- [32] G. E. P. Box and G. M. Jenkins. *Time series analysis, forecasting and control*. Holden-Day, San Francisco, 1970.
- [33] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer-Verlag, New York, 1987.
- [34] N. L. Gerr and S. Cambanis. Analysis of adaptive differential PCM of a stationary Gauss-Markov input. *IEEE Trans. Inform. Theory*, vol. IT-33, no. 3, pp. 350-359, May 1987.
- [35] A. Habibi. Comparison of n th-order DPCM encoder with linear transformations and block quantization techniques. *IEEE Trans. Commun. Tech.*, vol. COM-19, pp. 948-956, Dec. 1971.
- [36] E. J. Delp and O. R. Mitchell. The use of block truncation coding in DPCM image coding. *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-39, no. 4, pp. 967-971, April 1991.
- [37] A. N. Netravali F. W. Mounts and B. Prasada. Design of quantizers for real-time Hadamard-transform coding of pictures. *Bell Syst. Tech. J.*, vol. 56, no. 1, pp. 21-48, Jan. 1977.
- [38] H. J. Landau and D. Slepian. Some computer experiments in picture processing for bandwidth reduction. *Bell Syst. Tech. J.*, vol. 50, pp. 1525-1540, May-June 1971.
- [39] T. W. Huang. Digital picture coding. *Proc. Nat. Electron. Conf.*, pp. 793-797, 1966.
- [40] W. F. Schreider. Picture coding. *Proc. IEEE (Special Issue on Redundancy Reduction)*, vol. 55, pp. 320-330, Mar. 1967.
- [41] H. P. Kramer and M. V. Mathews. A linear coding for transmitting a set of correlated signals. *IRE Trans. Inform. Theory*, vol. IT-2, pp. 41-46, Sept. 1956.
- [42] T. Natarajan N. Ahmed and K. R. Rao. Discrete cosine transform. *IEEE Trans. Comput.*, vol. C-23, pp. 90-93, Jan. 1974.
- [43] J. Pearl. Basis-restricted transformations and performance measures for spectral representations. *IEEE Trans. Inform. Theory (corresp.)*, vol. IT-17, pp. 751-752, Nov. 1971.
- [44] A. Segall. Bit allocation and encoding for vector sources. *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 162-169, Mar. 1976.
- [45] T. R. Fischer and R. M. Dicharry. Vector quantizer design for Gaussian, gamma, and Laplacian sources. *IEEE Trans. Commun.*, vol. COM-32, no. 9, pp. 1065-1069, Sept. 1984.

- [46] P. F. Swaszek. A vector quantizer for the laplace source. *IEEE Trans. Inform. Theory*, vol. IT-37, no. 5, pp. 1355-1365, Sept. 1991.
- [47] T. R. Fischer. A pyramid vector quantizer. *IEEE Trans. Inform. Theory*, vol. COM-32, no. 4, pp. 568-583, July 1986.
- [48] N. S. Jayant P. Cummiskey and J. L. Flanagan. Adaptive quantization in differential PCM coding of speech. *Bell Syst. Tech. J.*, vol. 52, pp. 1105-1118, Sept. 1973.
- [49] M. Tasto and P. A. Wintz. Image coding by adaptive block quantization. *IEEE Trans. Commun. Tech.*, vol. COM-19, no. 6, pp. 957-971, Dec. 1971.
- [50] P. Noll. Adaptive quantizing in speech coding systems, *Proc 1974 IEEE Zürich Seminar on Digital Commun.* Mar. 12-15 1974.
- [51] M. Tasto and P. A. Wintz. A bound on the rate distortion function and application to images. *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 150-159, Jan. 1972.
- [52] V. Cuperman and A. Gersho. Vector predictive coding of speech at 16 kbits/s. *IEEE Trans. Commun.*, vol. COM-33, no. 7, pp. 685-696, July 1985.

BIOGRAPHICAL SKETCH

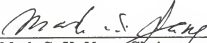
Fangshi Sun was born in Beijing, China, on April 20, 1951. After graduating from the 63rd Middle School of Beijing in 1969, he went to Inner Mongolia and worked there. He entered Baotou Iron and Steel Institute in 1979 majoring in electrical engineering and received his degree of bachelor in 1982.

He began his graduate study in the Department of Statistics, Purdue University in 1983 and received his degree of Master of Science in 1985.

He enrolled in the Department of Statistics, University of Florida in 1985 and expects to receive the degree of Doctor of Philosophy in May, 1992.

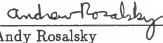
He was an instructor and taught several courses in mathematics and statistics before he came to the United States. At the same time, he did statistical consulting for the Baotou Iron and Steel Company. He has been a teaching and research assistant during his academic career at University of Florida.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



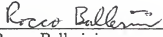
Mark C. K. Yang, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



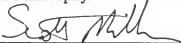
Andy Rosalsky
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Rocco Ballerini
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Scott Miller
Assistant Professor of Electrical Engineering

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1992

Dean, Graduate School